

52
267

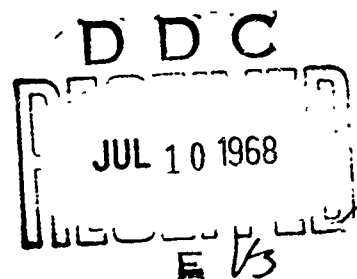
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
TECHNICAL INFORMATION PROGRAM*

AD 671269

TIP SYSTEM REPORT

October, 1967

M. M. Kessler



This document has been approved
for public release and sale; its
distribution is unlimited.

*Project TIP is supported by a grant from the National Science Foundation.
Project MAC is supported by the Advanced Research Projects Agency,
Department of Defense, under Office of Naval Research.

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

13

PREFACE

The Technical Information Program at MIT (TIP) is an experiment in information system design. Elements from articles in the physics journal literature are made available for on-line search and retrieval. The information is available from a network of consoles that have access to the MAC time-shared computer currently operational on the MIT campus.

The general configuration of the TIP system was described in a previous paper ("The MIT Technical Information Project", Physics Today, March, 1965). The present series of reports gives a more detailed and current description of the system based on the experience of the past two years.

These reports originated as a set of appendices in a proposal to the National Science Foundation. Because of their general interest, they are being reprinted and given limited circulation in the present format. We must apologize for the awkward pagination and for some confusion that may arise from the use of the appendix designation for internal reference between sections.

In the future, this document will be referred to as the "TIP System Report", October, 1967. The individual ~~sections~~ will be referred to by title and date.

CONTENTS

TIP SYSTEM APPLICATIONS	(APPENDIX B)
A description of TIP operations and a preliminary analysis of system experience.	
TIP PROGRAMS	(APPENDIX C)
A functional description of the programs available to the users of the TIP system.	
TIP USER'S MANUAL	(APPENDIX D)
A guide for on-line search and retrieval of the current literature in physics.	
TIP LIBRARY MAINTENANCE	(APPENDIX A)
A description of the methods and operating procedures used by TIP personnel in the production of a machine-usable library of journal articles, its maintenance and updating.	
A BIBLIOGRAPHIC SEARCH BY COMPUTER	(APPENDIX E)
The case history of a TIP user.	
BIBLIOGRAPHIC COUPLING EXTENDED IN TIME	(APPENDIX F)
Ten case histories.	
COMPARISON OF THE RESULTS OF BIBLIOGRAPHIC COUPLING AND ANALYTIC SUBJECT INDEXING	(APPENDIX G)
SOME VERY GENERAL DESIGN CONSIDERATIONS	(APPENDIX H)
A discussion of the goals and constraints that guided the evolution of the TIP system.	

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL INFORMATION PROGRAM*

TIP SYSTEM APPLICATIONS

A description of TIP operations and a
preliminary analysis of system experience.

M. M. Kessler

October, 1967

*Supported by a grant from the National Science Foundation

TIP SYSTEM APPLICATIONS

a. The On-Line User

Perhaps the most interesting and novel mode of engagement between a user and the TIP library is the direct on-line manipulation of the literature through the console keyboard. At the present time, there are some 200 consoles on the MIT campus, and a few off-campus, that can login to the MAC system. Each console may be used by any number of "programmers". The word "programmer" is used in a general sense to refer to anybody able and knowledgeable to warrant access to the system regardless of his intended use of the system. A "programmer" may be indeed a programmer in the standard sense, or he may be a scientist or librarian interested in using TIP only. Each of the several hundred programmers may use his number and pass-word on any of the consoles to gain access to MAC. A limited amount of time and memory space is assigned to each programmer who is then free to dispense it in any way that he considers best. Thus, although at the present experimental stage no money is involved in the user's consideration, an element of cost is nevertheless present in that the various functions available from the MAC system compete for the user's limited storage and processing allotments. If the user devotes x seconds to a literature search, he has that much less available that month for his other computer needs.

The time allotment is assigned in four shifts ranging in desirability from the eight-to-six daytime shift, a most convenient and therefore least available commodity, to the midnight-to-morning shift. Week-end time is also a designated shift. This arrangement promotes the distribution of work and encourages the use of a facility known as FIB (Foreground-Initiated Background). By using FIB, one may give a complicated set of instructions to the computer at some convenient time and instruct it to perform the actual work as a background problem, namely, at any time when the computer would otherwise remain idle. This facility is particularly useful when large amounts of print-outs are expected. Instead of the relatively slow letter-by-letter printing on the console at prime time, a fast line printer is used at some convenient time, and the material may be picked up the following morning. The memory allotment that is most critical to an application such as TIP is the number of records available on the disc. Each record accommodates 425 words. At the present

time, TIP has been allotted some 6000 records of which 1000 are used for various working purposes and the remaining 5000 for the TIP library. A more usual allotment of disc memory to an average user is perhaps 200 tracks, which is sufficient to store some 80,000 words.

The MAC system will accommodate 30 users as participants in the time-shared system. Fig. 14 shows a typical setting of a MAC console in an experimental laboratory. In this case, the console is surrounded by elements of an infrared interferometer and is normally used to make Fourier transforms of interferograms. The same console, however, may also be used to search the literature for this particular laboratory group (see Appendix E). Fig. 15* shows a console in the setting of the MIT Science Library. To the right of the console may be seen the microfilm library of the text corresponding to the TIP holdings and the finder-viewer-printer. A similar unit is located in a strategic location in the Physics Department and is available to graduate students working on their theses. A user interested in searching the TIP library may do so from his own console if he has access to one. If not, or if he wishes access to the microfilm library in addition to the console, he may come to one of the two public locations set up for this purpose. The instructions for using TIP are given in the TIP User's Manual (see Appendix D). The scheme of the manual is to get the user immediately involved with the system by means of a set of working examples that illustrate the various search and retrieval capabilities. The user need not be a programmer, nor does he have to be acquainted with the internal structure and organization of TIP. Questions are typed at the console; the answers come back immediately from the same console. A monitor program has been arranged that records the extent and nature of the user's participation in TIP. Fig. 16 is a six-month summary of the monitor record. Each point on the graph is the number of TIP engagements during a five-day period. On the average, some fifteen users ask for on-line use of TIP in each five-day period.

In a prototype system with limited facilities, it is not the number of users that is important, but rather their variety and distribution by type. No effort was made to urge or in any way "sell" TIP to a large population of users.

* Fig. 15 not included.

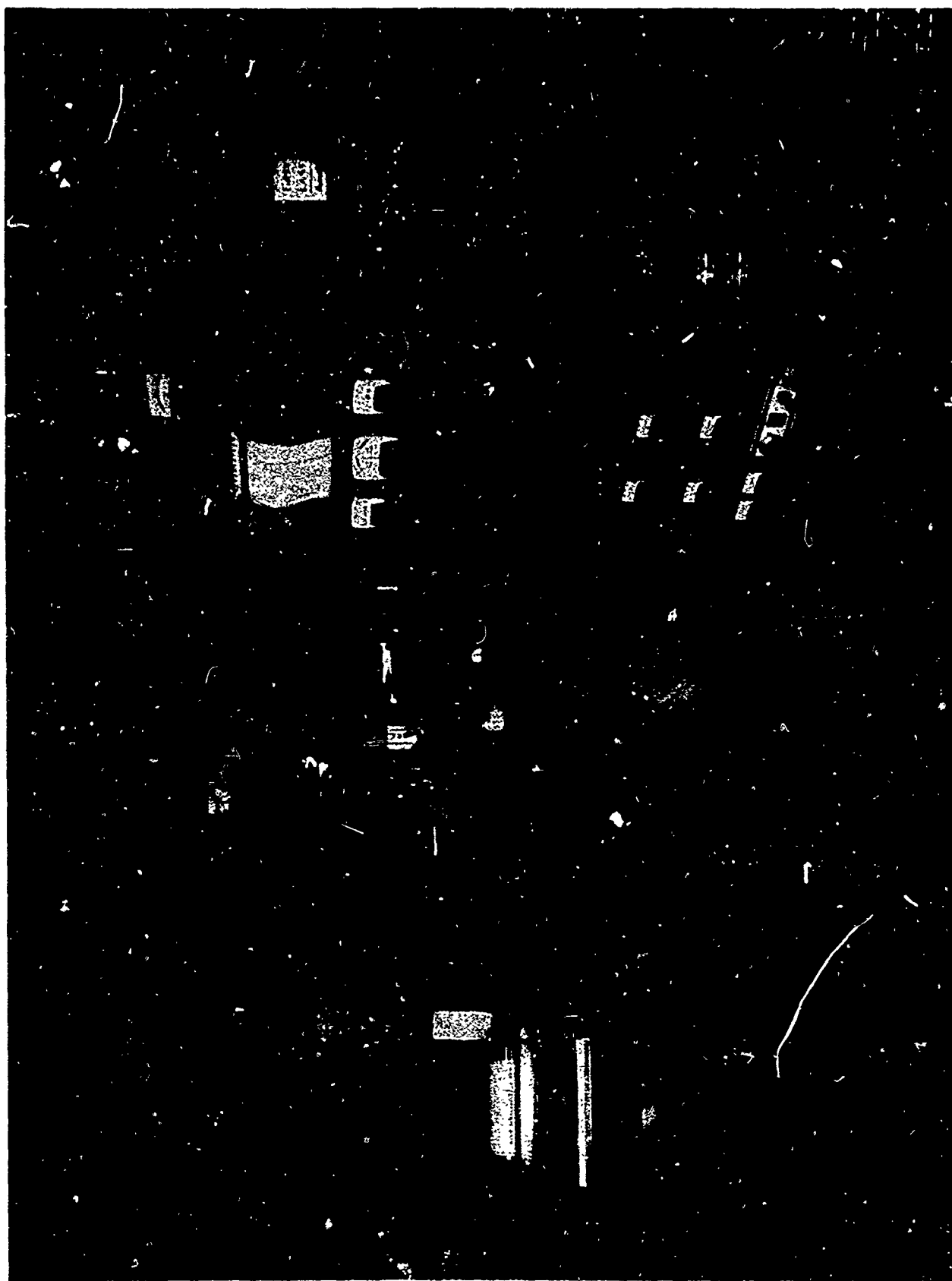


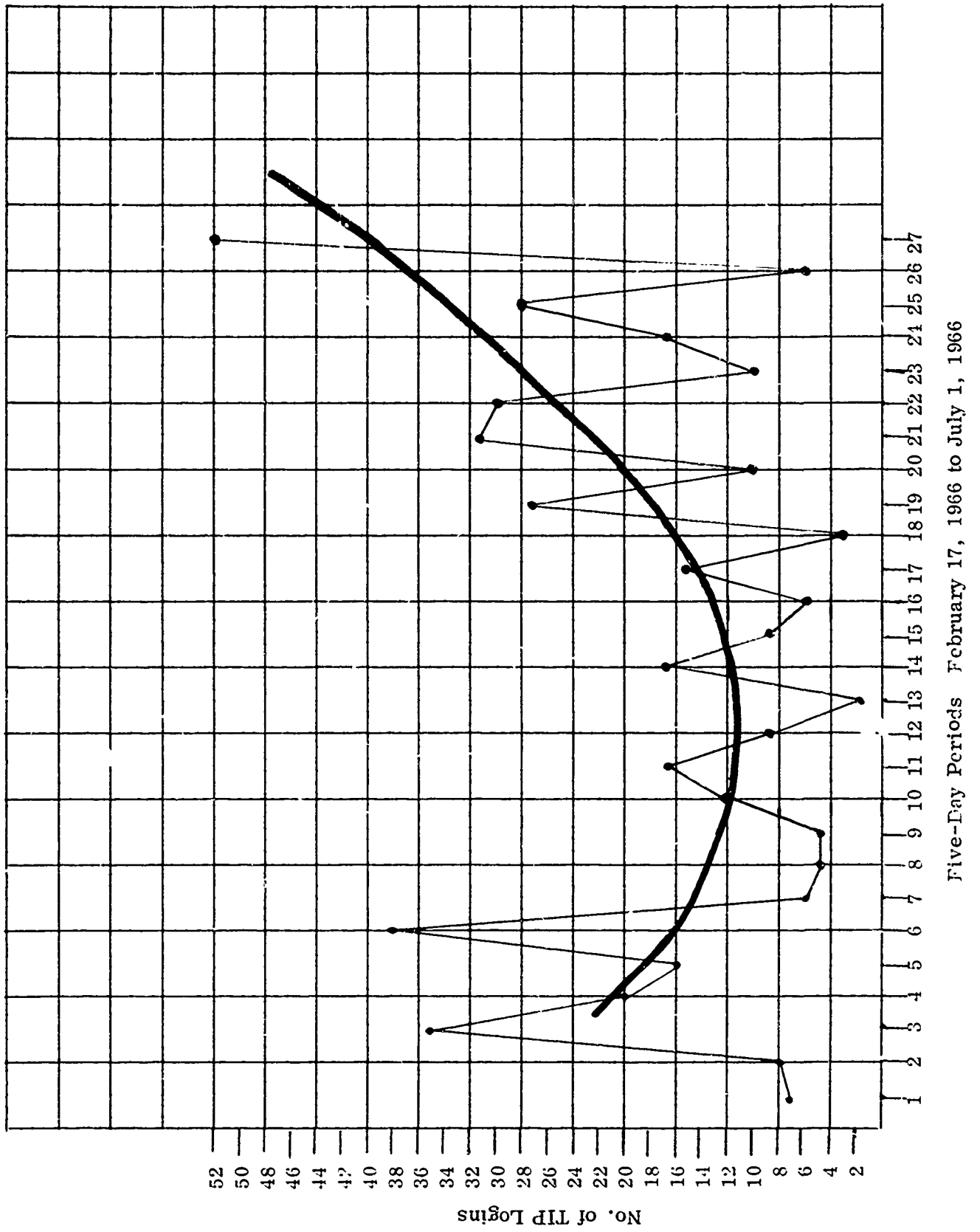
Fig. 14 MAC-TIP CONSOLE IN PHYSICS LABORATORY

Fig. 15 omitted from this report.

Fig. 15

Fig. 16

TIP LOGINS PER FIVE-DAY PERIODS



On the contrary, an effort was made to limit the number of users of a given type in order to accommodate as many different types as possible. At present, our users consist of students, research scientists, writers, librarians, historians, and sociologists of science as well as TIP and AIP personnel. The monitor record shows a wide diversity of engagement, ranging from the simple question concerning a title word or author to very extensive and sophisticated literature searches that make imaginative use of TIP (see Appendix E). The ready accessibility to a computer and all its facilities, its convenient presence in the office or in the laboratory, and the availability of a non-specialized interaction language are new and significant changes in the scientist's and student's environmental condition. The adaptation to this new factor must evolve through growth and accommodation of each to the other and cannot be imposed. A learning process must take place starting with the undergraduate student and proceeding through his graduate experience. Not only is the use of computers for literature search a new factor in research, but the very notion of using the literature at all must be reintroduced into the teaching process in a new and more effective manner. It is no exaggeration to say that the average engineer and scientist, to say nothing of the student, acquires as much information from an intelligent use of recorded knowledge as he does from direct experimentation in the laboratory. We recognize that the teaching of laboratory techniques is an important element in education, but the use of libraries and the scientific record has been largely neglected in the education of students. Indeed, it is frequently neglected even in the practice of mature and experienced scientists. The condition reflects a serious mismatch between the education of the scientist and the contemporary work of his seniors and masters. The mismatch is to be sought in the inefficient and undifferentiated flow of information through journal mailing lists (see Appendix E). Direct access by the scientist to the literature will bring about a closer relationship and will promote the use of the literature in a more personal and intimate setting. Evidence of this personal approach to the literature may be seen in the monitored TIP users' records where a wide spectrum of interests and needs are reflected in the variety of user requests and attempts to accommodate and bend the system to the personal habits and inclinations of the individual investigator. A sampling of

this variety will illustrate our point.¹

```
search phyrev new
find author not xyz
print identification, title, author
```

The user is asking for nothing more than a complete table of contents of the latest issue of The Physical Review in the TIP library. He is using the subterfuge of asking for all papers not written by XYZ. Since XYZ is not likely to occur as an author in The Physical Review, all papers are covered.

The user will soon discover that not all of the 38 journals are of interest to him. Thus, if he is interested in laser physics or solid state physics, he need not ordinarily consider journals specializing in nuclear physics. It is, at any rate, clear that as more journals are added to the TIP library, the command "search all" in any of its forms becomes impractical. Each user may compose a personal search command once and for all so that it need not be repeatedly typed at each session. Thus, he may enter,

```
search physical review new, physical review letters new, jetp new, and
jetp letters new
name saved file my library
```

```
In the future, the user will simply type,
read my library
```

and then give the find and output commands without having to repeat the list of journals each time. Indeed, the find and output commands may be incorporated into the read file if the user has an on-going interest in certain types of information.

It frequently happens that the results of a given search may be only the raw material on which further operations may be performed. In that case, one will want to save the output in TIP format. For example,

```
read my library
find title lasers
output save all
name saved file laser list
```

By saving the file, the user may perform subsequent searches for specific papers within the relatively short laser list and not have to search the entire TIP library for what he wants. The user will, however, learn from experience that the save command also costs him something because he thereby ties up storage

¹ See Appendix D, TIP User's Manual

space on the computer. He thus learns to evaluate his problem in terms of the trade-off between processing costs and storage costs. This trade-off is a significant consideration in the economic use of computers, and it is appropriate that the ultimate decision on that be left with users. It is possible to pre-process the information in various ways, to create indices and inverted files of all manner and generally to anticipate the reader's requests and store the answers in a format that will minimize the processing time and cost. This, however, is not always a worthwhile strategy since each redundancy of storage imposes a cost on the system. A proper balance must be found between processing costs and storage costs. In our present system, the data is stored once only in its original format corresponding to its schedule of publication. Each question generates its own answer as if it had never been asked before. Nor does the TIP system attempt to anticipate classes of questions and pre-process the data. For example, we do not store an author index or a citation index. The information is there, and the index is created each time a question is asked. This approach offers maximum economy in storage but bears the cost of repeated processing of the same data. There is no a priori reason why this principle should always be followed. It is simply the most convenient way for our operation at this time. If in the future we find that storage is cheaper and more available than a given amount of processing, we may well reformat the data to save processing time at the cost of additional storage. In the present system, the choice between repeated processing and additional storage is left to the user who is the ultimate consumer and, therefore, ought to decide the relative merits and costs of either approach. As of now, the almost universal choice of the users has been in favor of conserving storage space and paying the price of repeated processing. This one-sided response is no doubt due to the lack of familiarity with the system's capabilities as much as to a considered judgment of the economics. It is, nevertheless, our policy not to force the user in one or the other direction but to let each one evolve his own policy on the basis of his operational experience.

An analysis of the questions in the monitor record shows no particular trend or emphasis. The simplest search requests such as "find author 'a'" or

title word "t" were used most. More complex search schemes involving bibliographic coupling and list formation were used least as might be expected.

Somewhat more enlightening than the analysis of questions and requests is an examination of user types. The monitor records each time that TIP is engaged by a programmer number. No distinction can be made of users operating within a particular programmer number. Some of them may be individuals but, in view of the scarcity of tie-lines into MAC, it is common for a programmer number to be assigned to an entire group or laboratory and used by all members of the group.

Table A below shows the frequency of TIP logins by various programmers.

TABLE A
Frequency of TIP Logins by Various Users

<u>User Number</u>	<u>Number of Times Logged into TIP</u>
1	469
2	139
3	116
4	82
5	64
6	56
7	41
8	32
9	29
10 and 11	26
12 to 22	5-25
23 to 44	1-4

The data is based on a period of eight months and shows that altogether 44 programmer numbers logged into TIP 1205 times. Of the 44 programmer numbers who used TIP during this period, 22 used TIP 4 times or less. This is a heterogeneous class of users involving some who logged in merely for curiosity and others who have left MIT or otherwise did not find it necessary to use TIP. It does not exclude, however, some whose literature needs may be satisfied by such infrequent use. This group cannot be further analyzed because by the time they are classified into their various categories, the statistics are too meager to be meaningful. The category of heavier users may be examined in more detail.

The largest use of the system was by programmer number 1, who logged into TIP 469 times in the eight-month period under observation. This user came early to our attention, and his activities were well known to us. As a case history, it is most interesting because it represents the most sophisticated use of an on-line information retrieval system. Programmer number 1 is Professor Sanborn Brown and his group of co-workers in the Physics Department at MIT who used TIP to compile a book on basic plasma data that was published by the MIT Press². His experience with the TIP system was described in an article in Physics Today.³

Programmer number 2 (139 TIP logins) served Dr. Franz Alt and Mr. Russell Kirsch at the National Bureau of Standards who early in the history of our project acquired a MAC console located in their offices in Washington and used it primarily as an experiment in information retrieval in the NBS environment. Their results and experience have been documented and will be published.

Programmer number 3 (116 TIP logins) is the manager of the computer group for one of the large MIT laboratories, the Research Laboratory of Electronics. The individual to whom the number was assigned did not himself have any need to search the TIP literature, but his number was used by many members of the laboratory, ranging from students to professors. A check of this case disclosed that the group had several seminars that described TIP. As a result of this instruction, several students and staff members became proficient in TIP and used it frequently. Others asked questions of the operators who performed the search and gave the results to the inquirers. Still others had standing orders with the group or taught their secretaries to operate the system and perform periodic searches to their specification.

Programmer number 4 (82 TIP logins) represents a similar arrangement that evolved in the Nuclear Engineering Department. It is interesting to observe that in the last two cases, the computer group, rather than the librarians, acted

² Brown, Sanborn C. Basic Data of Plasma Physics 1966. Cambridge, Massachusetts: The M.I.T. Press, 1967.

³ Brown, Sanborn C., "A Bibliographic Search by Computer". Physics Today, 19, 5. (Appendix E)

as intermediary between the literature and the consumer.

Programmers 5 and 7 were individual members of the Research Laboratory of Electronics who used the system on their own number to supplement the service that they received from number 4.

An interesting relation developed between users 1, 5, and 7 and a professor who worked through number 4. These four scientists were all interested in plasma physics. By an informal agreement they used TIP-derived literature as a base to which they added their private non-TIP-maintained libraries of documents, reprints, books, etc. Through a division of labor, they covered different fields and would occasionally purge their lists of duplications. The private libraries of plasma physics were stored in TIP format and made available to all members of the group. The evolution of such user groups is an innovation that we are observing with great interest. Several programs and procedures have been worked out by TIP to encourage and facilitate this practice.

It is appropriate at this time to point out our attitude towards TIP users. As a general rule, we make no attempt to "sell" TIP or to urge its use. Our intent was to make its presence known through the usual channels used by MAC to inform its users and to let the user population grow at its own rate, based largely on the word of mouth propagation of the experience. Although we have from time to time discussed TIP with various users, we do not question them too much about their experiences. We do, however, observe the use of TIP as much as we can without disturbing the user's operations or otherwise introducing self-conscious, introspective responses on his part. When a certain need or behavior pattern appears in our monitor record, we try to modify the system in response to this feed-back. The modification will be designed to either encourage or discourage the observed behavior, depending on our judgment of its worth. For example, when we saw that several users were interested in selecting their own list of journals to be scanned rather than the entire TIP library, our response was to introduce the READ facility previously described. Similarly, the need of the users in groups 1, 4, 5, and 7 to compare lists and eliminate duplications started us on an entire series of programs designed to

facilitate list handling.

Programmer number 6 (56 TIP logins) was a manager of the document room at Project MAC who tried to set up a file of MAC memoranda and internal reports for the use of the MAC community.

The rest of the programmers were individual scientists who used the system intermittently.

It is too early and the experience too scarce to make any generalizations with regard to the acceptance of an on-line information system by the community of scientists. Our record does not offer a strong enough statistical base for any conclusions. The case histories that we discussed show trends that will be further observed and exploited. We feel that the prototype nature of the system must at all costs be maintained in order to conserve the possibility for change and evolution. Several tentative remarks can, nevertheless, be made even on the basis of the very limited experience that we have so far had with TIP on-line users.

Those working in the field of information retrieval must never forget that what to them is a full-time professional occupation is in many cases of only peripheral interest to the working scientist. We have not observed a stampede of scientists clamoring for the use of TIP. This may, of course, be due to the fact that TIP offers only a trivial service. However, the fact that a few of our scientific colleagues did find it worthwhile to make exhaustive use of the system leads us to believe that the primitive nature of TIP is not the only reason for its modest use by the MIT physics community. The novelty of the engagement and the general neglect of literature search activity as a significant occupation for scientists may account for at least some of the resistance to its use. As a general rule, the use of TIP depends on the availability and convenience of its machinery. Whether TIP is used at all and to what extent depends not only on its intrinsic worth as tested by various criteria in the laboratory but also on the local computer environment. At the present time, MAC has some 200 input-output consoles. Of these, perhaps not more than 30 are located within easy reach of those whose literature needs are matched by the TIP library. Furthermore, the limited capacity and high utilization

factor of MAC put a strain on the system, particularly during the most desirable daytime hours. The use of TIP, therefore, involves certain delays and frustrations that have nothing to do with its designed capabilities but are due to the local conditions of computer operations. All kinds of tests and evaluations of TIP have been made under controlled conditions. However, in observing its use in a working and realistic environment, it is most difficult to separate those factors that are intrinsic to TIP from those that are incidentally imposed on it by local engineering and economic factors. It is, nevertheless, the totality of the phenomenon that must be observed and, as such, it cannot be separated from its environment. It would not suffice to develop a retrieval system of the most marvelous sensitivity if the technological environment were not prepared to support it. Our tendency is to consider TIP capabilities as inseparable from the computer and communications systems on which it depends and direct the growth and evolution of one in such a way as to fully exploit the other. With the above considerations in mind, we can say even now, with our limited experience, that given a convenient access to an on-line system such as TIP, the community of scientists will slowly develop the habits of making intelligent use of it.

b. Non-Interrogated System Response

Our concept of the proper relation between the published literature, retrieval, communication, and the population of readers is shown in Fig. 17. In this chart, retrieval and selection are indicated as elements of the processing scheme that the published literature must undergo prior to its proper dissemination to the scientific community. The processing function operates on the totality of the published literature and regroups the papers into meaningful bundles which may then be addressed and delivered to the appropriate readers. The output from processing takes two channels, i.e. interrogated output when the address is an integral part of the interrogation, and non-interrogated output when the address is part of the processing function. The non-interrogated output may be broadcast or published through a variety of media to the community at large, or it may be transmitted to specific persons or groups. The interrogated output is normally available to the interrogator only, although it may at times be considered general enough to be worthy of broadcast, especially if the original interrogator adds to its value in some way by rearranging or otherwise enriching its content. The arrows leading from the bottom of Fig. 17 back to the input stage indicate a feed-back loop from the user.

Referring to Fig. 17, the reader will recognize that on-line use of TIP as described above is an example of interrogated response communicated to the point of origin, namely the user at the console. This is still essentially true even when the response is delayed, or when it is triggered by a standing READ order or even when the results are made available to others. The point is that in this type of engagement the question is initiated not by the system and not by some intermediary but directly by a user-interrogator who, initially at least, is also the sole consumer.

We shall now present an example of a different mode of operation. In this case no user interrogation is involved. The information within the system and the logical machinery at its disposal generate a message with a proper address. The intended user receives the message without any initiative on his

TECHNICAL INFORMATION FLOW DIAGRAM

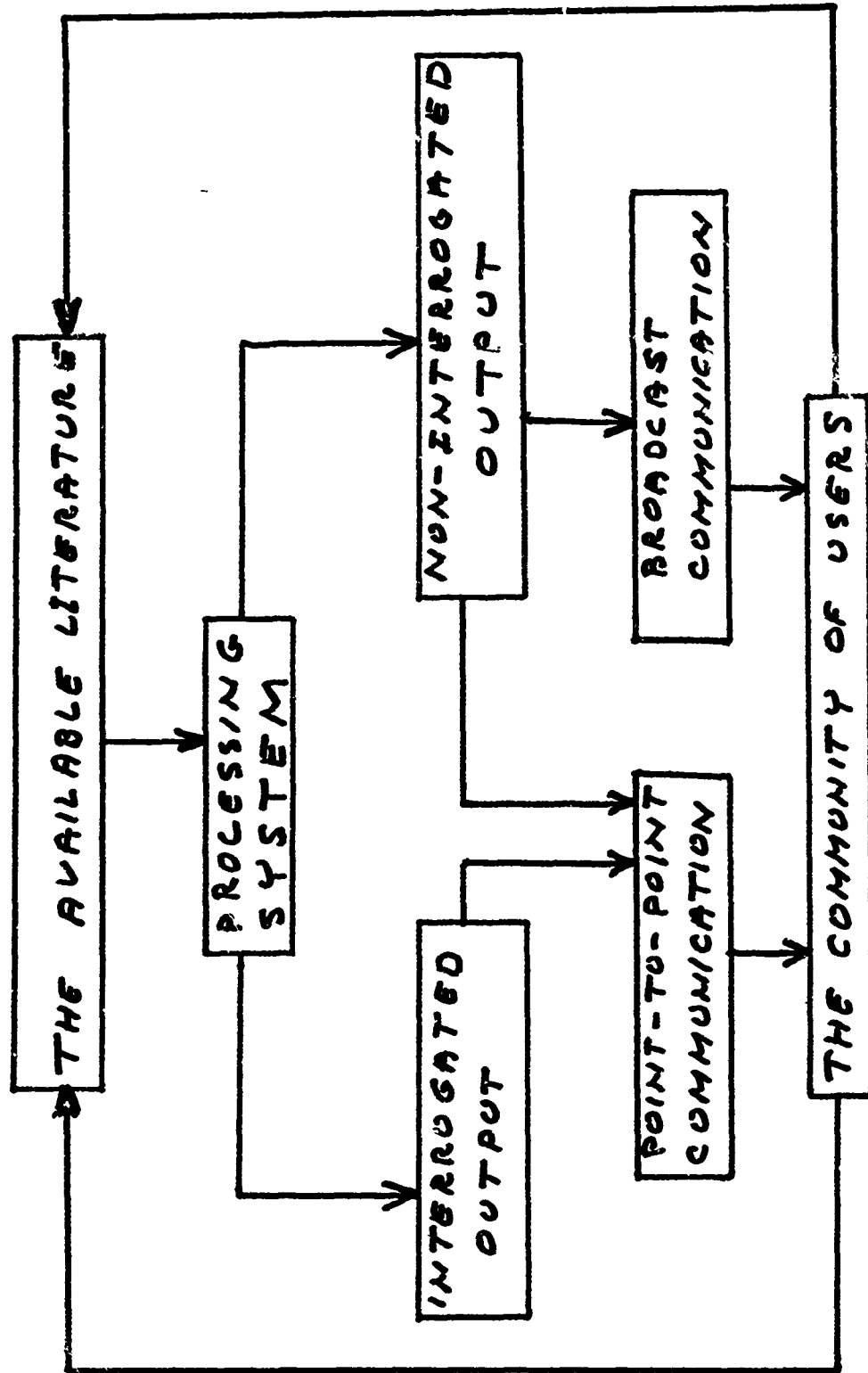


FIG. 17

part being necessary -- indeed, at times even without his knowledge of why he was chosen to receive the information. No contact or acquaintance between the recipient and Project TIP is necessary since all decisions are internal to TIP. This will be a case of non-interrogated, point-to-point communication.

The TIP library of the past three years contains the authors, titles, and locations of some 30,000 papers in the physics literature. Among them is a large number of papers by MIT authors. We create a file of all MIT authors and their papers. To this we add their local addresses at MIT. Each week as new literature is entered into TIP, it is searched for references to the MIT-authored papers. As such cases are found, they are stored, and subsequently a message is printed such as in Fig. 18, informing the author where, when, and by whom his paper was cited. The message is sent to the proper party by interdepartmental mail.

Fig. 19 is a flow chart of the process that generates this message unit. As new literature arrives each week, it is processed into TIP format and will eventually join the TIP public library. However, its status as NEW LITERATURE is maintained for a week or so. Before losing its identity as NEW LITERATURE, we extract all articles by MIT authors. This information is checked against the local directory for correct local address and to remove those who are no longer at MIT. This constitutes the weekly addition to the master list of MIT authors. Each time a new MIT directory is published, the master list is itself checked in order to up-date the local address and to remove non-MIT authors. The NEW LITERATURE is again searched and matched against the master index of MIT authors and their publications. Each time a reference is found to a member of the master list, the proper communication is printed and addressed.

Fig. 18 is an example of an alerting message, non-interrogated and precisely addressed. The two components of the message, namely its content and its address, are both generated by the internal information and logic of the system. Both contain possibilities for limited expansion. The list of recipients could be expanded to all authors in the TIP library but not to those who are not in the TIP library but might find the message useful. Even the limited expansion to all authors of TIP papers involves a serious technical

Project TIP
Room 14S-310

INTERDEPARTMENTAL MAIL

Mr. Bennet Brabson
575 T. S. 4th Floor
MIT

Your paper published in:

PHYSICAL REVIEW LETTERS V13 P558
TOTAL CROSS SECTIONS AND ANGULAR DISTRIBUTIONS FOR P-R CHARGE
EXCHANGE IN THE SECOND AND THIRD RESONANCE REGIONS

Has been cited in:

IL NUOVO CIMENTO V45 P983
A PHOTOPRODUCTION CROSS-SECTION FOR INCIDENT PHOTON ENERGIES
FROM 800 TO 1000 MEV

BACCI C.
PENSO G.
SALVINI G.
MENCUCCINI C.
SILVESTRINI V.

Fig. 18

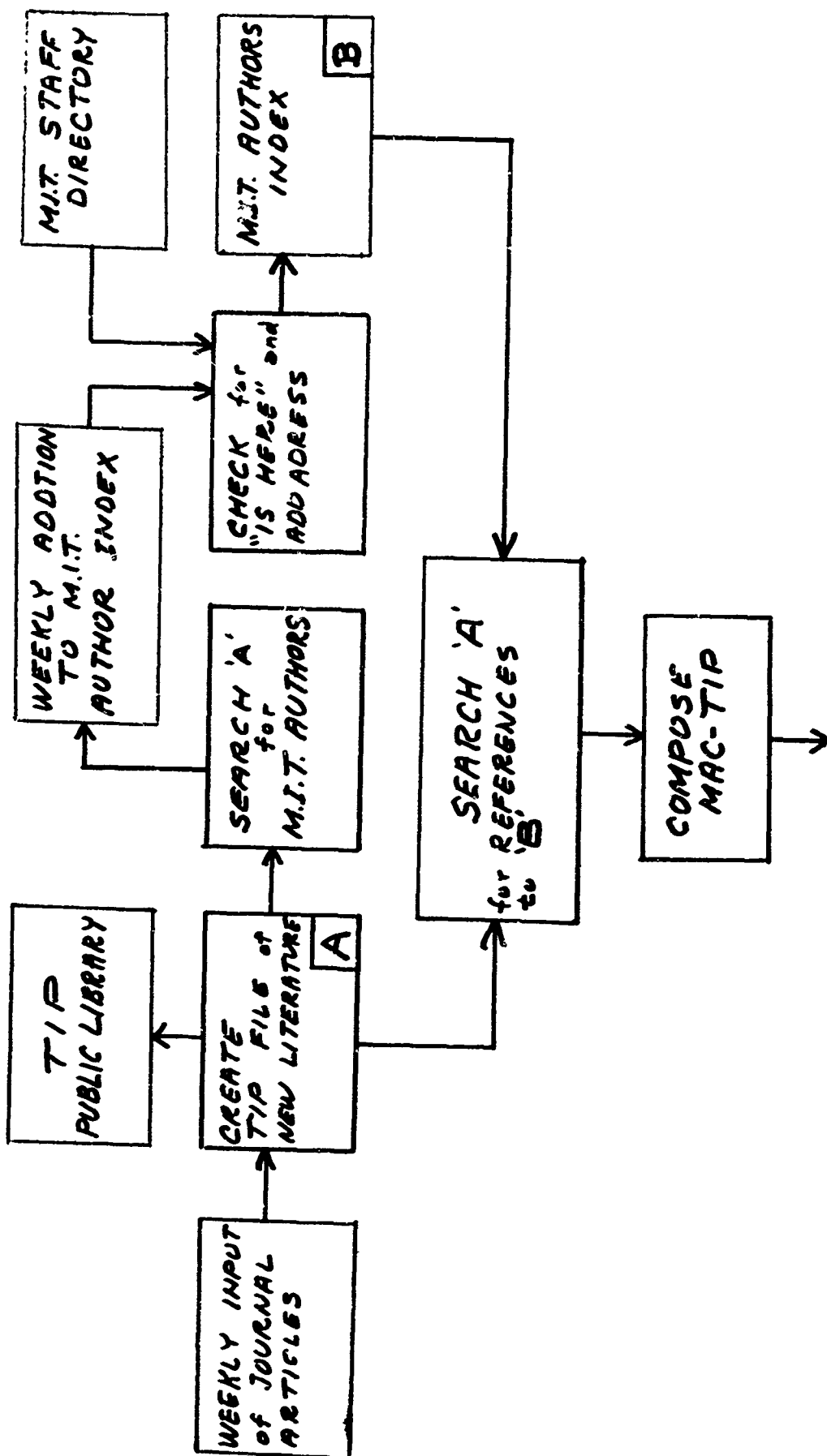


Fig. 19

difficulty in the maintenance of a sufficiently current address list without the recipient's cooperation. If the recipient is to be required to inform the system of changes in his address, then, of course, the scheme is no longer to be classed as uninterrogated. Without the recipient's cooperation, the problem of proper address is formidable.

Extension of the scheme to include information other than the act of citation is quite promising. The fact that an author's paper has been cited is used as a criterion of likeness that makes it probable that the citing paper will be of interest to the recipient. Other criteria of likeness have been tried and tested and will be used as message generators. Similarities of title between papers in the weekly TIP input and those of the recipients can be used as an indicator of likeness. Of particular significance and promise in this connection is the criterion of bibliographic coupling.^{4, 5}

⁴ Kessler, M. M., "Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing". American Documentation, 16, 3, July, 1965, pp. 223-233. (Appendix G)

⁵ Kessler, M. M., "Bibliographic Coupling Extended in Time: Ten Case Histories". Inform. Stor. Retr., 1, pp. 169-187. (Appendix H)

c. System-Originated Public Information

Referring again to Fig. 17, we shall now treat an example of a non-interrogated, broadcast or widely distributed mode of TIP operation. In developing the strategy for this mode, the aim was to examine a large body of literature, say all TIP-1966, and to sort it into several subgroups such that each of them will be of interest to a large community of physicists. For example, remove all papers from TIP-1966 of probable interest to plasma physicists or laser physicists. These lists, which may consist of several hundred papers, are not addressed to any one single recipient but to large classes of users. Typically, such lists may be published on a regular and continuous basis, say every year. The value of such massive lists of papers in broad fields such as plasma or laser physics is much enhanced if some internal arrangement and organization can be built into the strategy. We shall call such lists bibliographies and present by way of illustration the strategy and TIP operations that produced a 1966 bibliography of the laser literature.

The problem simply stated was to examine the TIP literature of 1966 and to use any combination of processing schemes at our disposal in order to extract a group of papers that are most likely to be of interest to laser scientists. These papers were to be arranged and presented in such a way as to be of maximum use to the intended reader. At no stage in the process must we require expert judgment and knowledge of the science involved. This judgment is to be deferred to the very last stage, when the intended user reads the results of the TIP output. Each user will then exert his own judgment and select his own material. The TIP contribution is merely to reduce drastically the number of papers from which selection is to be made.

The appreciation of our goal and intent is important in evaluating the results. The 1966 data in the TIP library consists of 15,000 papers from 32 physics journals. Included in this group is a smaller set likely to be of interest to laser scientists. All that is necessary to retrieve this smaller subset from the larger set of papers is to line up enough laser physicists and ask them to make the decision on each paper: Is it or is it not likely to be of interest to your colleagues? This scheme could be used except for the fact that there are

not enough laser scientists willing to perform the task. The job is, therefore, given to groups known variously as indexers, literature experts, key-word-assigners, etc., etc. who perform the task with various and, at times, questionable degrees of success. The goal we set for TIP is not to find "all papers that have to do with lasers", but rather to form a group of papers, much smaller in number than the total literature, such that the probability is very high that laser papers are within it. Individual laser scientists will still have to examine each paper and decide, "Is it or is it not of interest to me?", but the number of papers so examined should be much smaller than the original library of 15,000 items. The second requirement of the process is that there be a very high probability that laser-related papers do actually find their way into the smaller group.

Fig. 20 is an example of the search strategy that we employed to produce the group of laser papers. Several steps are involved.

1. Examine the TIP literature of 1966 and find all papers that have the word "laser" in the title. The results of this search are printed in List A and are also saved in the computer memory as "file a" for subsequent use.
2. Examine the TIP literature of 1966 and find all citations to papers in "file a". This is printed as List C and also saved in the computer as "file c".
3. Examine all the references of the papers in "file a" and arrange in order of descending frequency of occurrence. Call this "file b".
4. Find all members of "file b" that are also in TIP-1966. Print these as List D and store as "file d".
5. Remove the most frequently cited papers in "file b" (say, the top 10%) and store these as "file f".
6. Examine TIP-1966 and find all citations to papers in "file f" that do not contain the word LASER in the title. Print this as List E and store as "file e".
7. Examine all the titles of the papers in "file e" and arrange the title words by descending frequency of occurrence. (In this program,

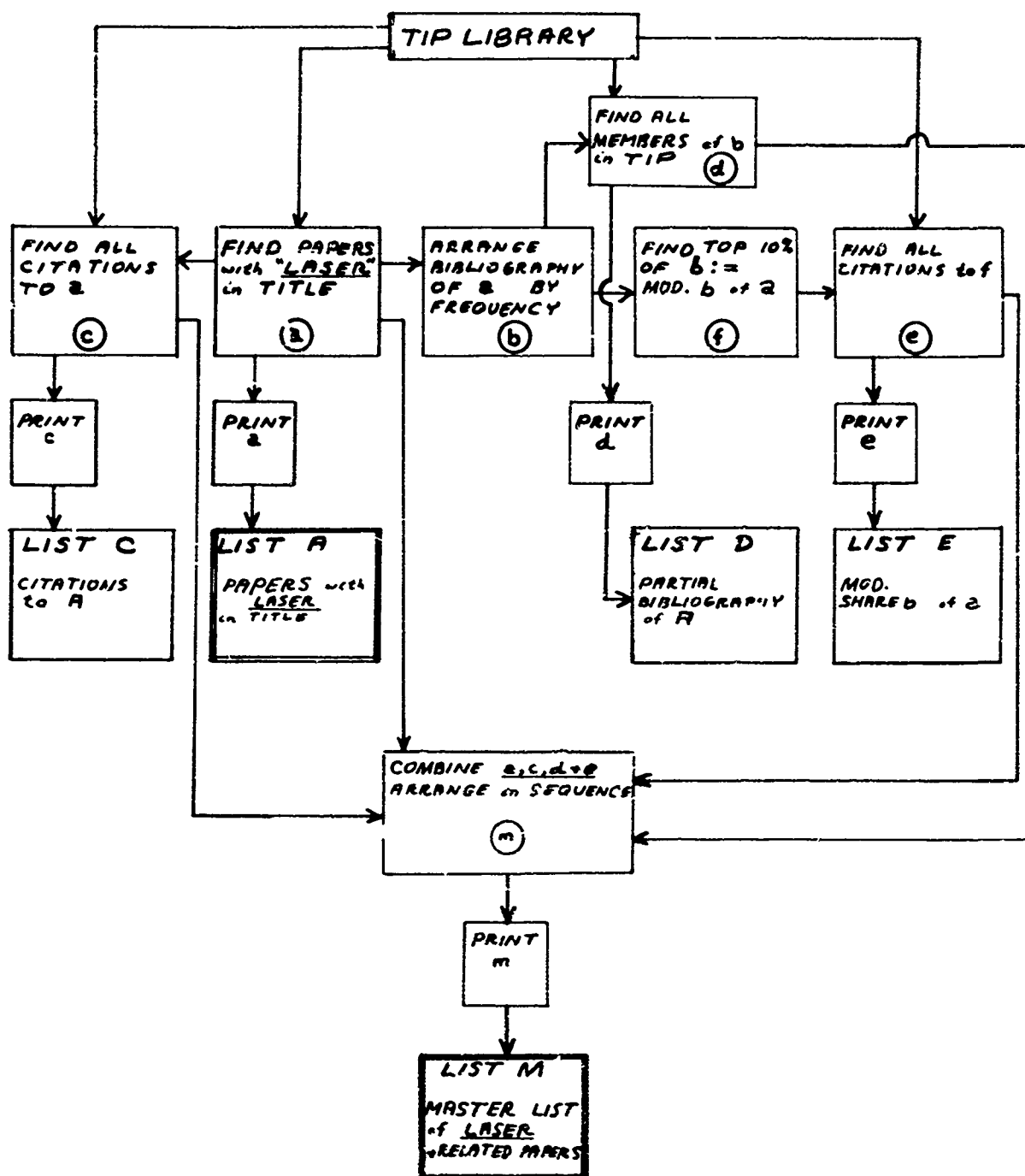


Fig. 20

there is a small list of words that are ignored in the count such as "of", "the", "by", etc.) Remove the most frequently occurring words that are judged pertinent and call this "file g". (This is the only point in the process where some human judgment comes into play.)

8. Examine TIP-1966 and find all titles that contain the words in "file g". Print as List H and store as "file h".
9. Repeat steps 2 to 6 and reiterate if necessary.
10. Combine files "a", "c", "d", "e", and "h", eliminate duplications and call it "master file m". Print as Master List M.

END of document retrieval process. Document arrangement will be discussed later in this section. Except for the single detail in step 7 and, to some extent, in step 5, the entire process consisting of steps 1 to 10 is automatically performed by the computer using publicly available TIP programs. Indeed, the entire operation could be grouped under one command and performed automatically on the literature given the few parameters necessary such as literature range (TIP-1966), starting word (LASER), and the cut-off points in steps 5 and 7 (say, top 10%). This would come as close as we can presently get to an entirely automatic bibliography-making machine.

It is good to review steps 1 to 10 to see what was accomplished in more familiar terms. There is no question about the results of step 1. If an author sees fit to put the word LASER in a title and if this is agreeable to the editors and referees, the pertinence of List A is beyond dispute. This list is, therefore, a definitive starter for the entire process. In step 2, we generate a limited citation index to the papers in List A. It is limited in the sense of containing only those citations to List A that were published in TIP-1966 papers. This, of course, is consistent with our goal to generate a LASER bibliography that is a subset of TIP-1966. Since papers containing the word LASER in the title are of prime and unquestioned pertinence to us, we make the assumption that what these papers cite, being but one generation removed,

is also of interest. These are, therefore, filtered through the TIP-1966 literature range and retrieved. So far, we have retrieved the definitive LASER-containing list and the TIP-contained papers that either cite them or are cited by them. In steps 5 and 6, we perform a modified share b function on a fictitious bibliography that consists of papers most frequently cited by those containing LASER in the title. The injunction to eliminate those that contain LASER in the title merely removes a redundancy nuisance, since the latter had already been discovered in step 1. Steps 7 and 8 involve a subtle maneuver. Clearly, the word LASER is not the only title word of interest that one ought to look for. There may be synonyms, equivalents, or otherwise related title words that could serve as significant search criteria. It is not likely that a title that contains the word LASER will also contain its synonym or equivalent. We look for such words in "file e". The question we ask is as follows. "File e" contains papers that do not have LASER in the title. They are, however, related to the laser subject because they cite those papers that are most frequently cited by LASER-containing titles. If, then, "file e" has no LASER in the title, what words are frequent and prominent in it? The presumption is that, at least as far as retrieval is concerned, these words are to some extent equivalent to LASER. In an actual experiment the words so discovered were "optical maser", "brillouin", "stimulated emission", and "coherence". These words would have come to mind to any experienced laser scientist who could have included them in step 1 so that steps 7 and 8 could be eliminated altogether. It is interesting, however, that the same judgment may be imitated by having the computer perform steps 7 and 8. Steps 9 and 10 are optional and merely close the process.

The retrieval of a list of several hundred papers, even when they clearly relate to a subject such as LASERS cannot be considered the end product. Some manner of grouping, ordering and arrangement has to be provided in order to facilitate its use by the interested scientist. The scheme of presentation that we follow derives naturally from our scheme of retrieval. It consists of lists, indexes, and groups. A primary list of papers is found

according to some criterion. In our example, the criterion was the presence of certain words in the title (laser, maser). To this primary list of titles we added two additional lists that derive from the first by virtue of citing it or being cited by it. A third group of papers is added to these by virtue of bibliographic coupling. These three lists consist of titles, authors, and journal-volume-page identification of the papers. For simplicity of reference, each paper is given a sequence number.

The above papers are then fused into one file from which we produce three indices.

1. The author index consists of an alphabetic arrangement of all the authors followed by the sequence numbers of each author's papers.
2. The word index consists of an alphabetic arrangement of the title words followed by the sequence number of the papers that contain those words.
3. The citation index consists of a serial arrangement of all references and the sequence number of the papers that make the reference.

Finally, the papers are clumped into groups, G_A and G_B , so that the choice of a given paper suggests others that may be of interest. The groups are made up according to the carefully defined criteria for bibliographic coupling.

- a. A single item of reference used by two papers is called one unit of coupling between them.
- b. The coupling strength between two papers is measured by the number of coupling units (n) between them.
- c. A number of papers constitute a related group G_A if each member of the group has at least one coupling unit to a given test paper P_0 .
- d. A number of papers constitute a related group G_B if each member of the group has at least one coupling unit to at least one other member of the group.
- e. Definitions of G_A and G_B may be modified by assigning higher values to " n ".

The total output of this process is arranged as follows:

1. The primary list
2. Those that cite 1
3. Those that are cited by 1
4. Those that are related to 1 by bibliographic coupling
5. Author index
6. Word index
7. Citation index
8. G_A grouping
9. G_B grouping

As a matter of possible interest to the user, we also include:

10. Frequency distribution of papers in lists 1 - 4 among the various journals

The search strategy represented by steps 1 to 10 is only one of the several variations that can be constructed with TIP facilities. The results vary in efficiency, depending on the literature and purpose. It is not always possible to start with a word-in-title search. Not all seemingly possible words are suitable for searching. For example, the word "nuclear" or "particle" may not be as productive a search word as "laser". There are, however, other starting possibilities and other search strategies that may be tried. The result is always a list of papers retrieved from the TIP library that we presume is of interest to a class of scientists. The validity of this process is subject to test and judgment by knowledgeable people. Various testing procedures have been tried. We shall briefly describe one such test on a TIP operation.

In April, 1966, the American Institute of Physics interested Mr. E. V. Ashburn to examine the results of a TIP search. For several years in the past, Mr. Ashburn had been interested in collecting and publishing a bibliography on lasers. He compiles the list by using the standard literature search techniques such as combing the abstracting journals, tables of contents, references, etc., etc. It is a painstaking and laborious job that Mr. Ashburn performs as a labor of love. When he came to us in April, 1966, he had finished his compilation of the 1964 laser literature which had already been published. It was decided to repeat this work on TIP and have Mr. Ashburn evaluate the results. The search strategy that we used was somewhat simpler

than the one described above and consisted of three steps.^{6, 7}

1. Search 1964 TIP library and find all titles containing one of the following words: laser, maser, brillouin, stimulated emission, and coherence.
2. List all references occurring twice or more in papers of step 1.
3. Search TIP-1964 for articles that refer to list 2, but eliminate the search words of list 1.

The combined results of operations 1 to 3 consisted of 321 articles. According to Mr. Ashburn's judgment, 62 of these were irrelevant to the subject of lasers, leaving 250 relevant articles. Mr. Ashburn's previous manually produced bibliography contained 280 articles. When these two lists were compared, again according to Mr. Ashburn's judgment, his bibliography had 59 papers that TIP failed to locate, while the TIP list included 41 papers that were not found by Mr. Ashburn.

⁶ Ashburn, Edward V. and Bela A. Lengyel, "Bibliography of the Open Literature of Lasers. VI". Journal of the Optical Society of America, January, 1967, pp. 119-146.

⁷ Laser Literature, Edward V. Ashburn, editor and compiler. North Hollywood, California: Western Periodicals Company, 1967.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL INFORMATION PROGRAM*

TIP PROGRAMS

A functional description of the programs
available to the users of the TIP system.

M. M. Kessler
October, 1967

*Supported by a grant from the National Science Foundation

TIP PROGRAMS

The TIP project, being a client of the larger MAC system, makes use of all its programs and facilities as well as those designed specifically by us for TIP purposes. The programs available to us fall into six classes.

- a. Programs developed by MAC that maintain the overall CTSS system. These we simply accept and take for granted as part of our working environment.
- b. Facilities developed by MAC that offer aid in writing, compiling, and debugging programs. These are primarily of interest to programmers and represent a very basic assistance from MAC.
- c. Programs developed by MAC for non-programming purposes and available to the public. The example of this most useful in the TIP context are a group of programs that enable the user to edit, format and compose text.
- d. Programs developed by TIP staff for maintenance and internal processing. These are not of interest to the user of TIP.
- e. Programs developed by TIP staff for the specific purpose of literature search and retrieval. These will be the main subject of this report.
- f. Programs developed by TIP users for special and private purposes. These are not usually available to people other than the originators and are, indeed, frequently not known to us at TIP.

The programs in groups a, b, and c are fully described in the various MAC publications.¹ Program Group "d" will be discussed in another section of this report.² At this point, we shall elaborate on Group "e": programs developed by TIP for the express purpose of literature search and retrieval.

1. The Search-Find Package. The computer operates on the TIP

¹ The Compatible Time-Sharing System. A Programmer's Guide. Cambridge, Massachusetts: The M.I.T. Press, 1965.

² "TIP Library Maintenance", Appendix A.

library (2) by means of a set of programmed commands. The command language is the interface between the user and the literature on which he operates. It is, in a way, the only aspect of the computer that a user need be aware of. As such, it is at the very heart of the system and most sensitive to human needs. One must make the distinction between what is called the programming language and the language of engagement between man and machine. The programming language is a set of instructions to the computer written in the esoteric vernacular that matches computer capabilities. It is largely incomprehensible when regarded as English text. The language that a user would comfortably wish to use when performing a literature search should be as close as possible to natural English and should have a sufficient vocabulary of verbs and nouns to allow a flexible group of choices to be made by the user without requiring the intermediary of a programming expert. In our case, the programming language is largely FAP and need not concern us here. The language of engagement has evolved through several generations of use and is constantly being changed and adjusted in response to user feedback.³

To engage the MAC system, one dials the proper number on the dataphone attached to the keyboard machine. When contact with the computer is made, a characteristic dial tone is heard. There follows a login procedure involving a password. This identifies the caller as a legitimate user of the system. If a free line is available, he will be so informed and may then proceed to use the computer in any of its capacities. If the library function is desired, the user will type the word TIP. This instruction makes the computer receptive to the TIP command language. The simplest form of engagement may be initiated by specifying three commands:

search: defines literature range to be examined.
find: states the items to be found.
output: defines the content of the desired output.

³ TIP User's Manual, Appendix D.

The "search" command may have the following elaborations:

search: all - will search the entire literature in store.
search: all new - will search the last volume of each journal.
search: phyrev all - will search everything in store of a given journal
(in this case, The Physical Review).
search: phyrev v. 120 to v. 135.
search: phyrev v. 120.

The program will search as specified above and detect any item described under the "find" command which itself has a variety of possibilities.

find author: Smith - will find all papers that include Smith among its authors.
find title: cryogenics - will find all papers that contain the word "cryogenics" in the title.

One may similarly define a location.

find location: Massachusetts Institute of Technology,

or a citation

find citation: 1 131 1165 - will find all papers that cite the article in The Physical Review (code 1), volume 131, page 1165.

The "output" command selects one or more of the various options available for output. One may call for immediate print, or one may wish a preliminary count of the items found before asking for print. One may also store the results under a file name and call for it later for further manipulation if immediate response is not required.

The above commands may be given in any order and any combination. Since the bulk of the computer time involved is spent on reading the material, it is frequently convenient to combine many requests and find them all in one search of the literature. Fig. 1 shows one such example. Note that two words, "a & b", typed on one line under "find" are interpreted as "a and b", while two words on two lines are interpreted as "a or b". Thus,

find author: Smith Jones

means find papers by Smith and Jones.

find author: Smith
find author: Jones

COMPUTER RESPONSE TO MULTIPLE QUERY

NOTE: Lower case letters are instructions from the user.
Upper case letters are computer response.

tip
W 1050.9

search phyrev volume 133
find
title= multimode cavities
author= garbacz robert j.
location= valparaiso university
citation= 110 109 345
output
print title author location citation
go

PHYSICAL REVIEW

VOLUME: 133

PAGE: 0014

BISTATIC SCATTERING FROM A CLASS OF LOSSY DIELECTRIC SPHERES
WITH SURFACE IMPEDANCE BOUNDARY CONDITIONS

GARBACZ ROBERT J.
COLUMBUS, OHIO
THE OHIO STATE UNIVERSITY
DEPARTMENT OF ELECTRICAL ENGINEERING

PAGE: 0069

GREEN'S FUNCTION THEORY OF MULTIMODE CAVITIES

KEMENY G.
LEXINGTON, MASSACHUSETTS
KENNECOTT COPPER CORPORATION
LEDGEMONT LABORATORY

PAGE: 0165

DEUTERON INTRABOND MOTION AND FERROELECTRICITY IN KD2PO4

SILSBEE HENRY B.
UEHLING EDWIN A.
SEATTLE, WASHINGTON
UNIVERSITY OF WASHINGTON
SCHMIDT V. HUGO
VALPARAISO, INDIANA
VALPARAISO UNIVERSITY

PAGE: 0253

THERMAL CONDUCTIVITY AND PHONON SCATTERING BY MAGNETIC
IMPURITIES IN CdTe

SLACK GLEN A.
GALGINAITIS S.
SCHENECTADY, NEW YORK
GENERAL ELECTRIC RESEARCH LABORATORY
J0110 V109 P0345

SEARCH COMPLETED

Fig. 1

means find papers by Smith or Jones. This convention is also true for words in the title and other items. It is also possible to use the exclusion command, "but not". For example,

find title: microwaves but not spectroscopy.

In Fig. 1, the search range was vol. 133 of The Physical Review. The system is requested to find papers that contain the phrase "multimode cavities" or those by a given author or coming from the named university or if they happen to cite a given reference. Each of the defining characteristics could be expanded into a list and combined in any logical combination of "and", "or", "but not". The computer output is printed in capital letters. The paper on page 0069 was found because the word "multimode" and the word "cavities" were in the title. This paper would have been retrieved even if the two words did not follow each other in the title. If the title had been "Multimode Vibrations in Non-Resonant Cavities", it would have been retrieved. If the exact sequence of words is desired, one may use (*) as an implicit blank. Thus,

find title: multimode * cavities

would find only the two words as shown. There are other conventions that concern the beginnings and endings of words. Thus,

find title: superconduct

will find titles containing the word "superconduct" as well as "superconductor" and "superconducting". But if we request

find title: superconduct *

will find only those titles that contain the word as shown. Similarly,

find title: conduct *

will behave as above, whereas

find title: +conduct *

will find also "superconduct" or any other prefix joined to "conduct", but not any suffix to it since the (*) blocks that.

As a general rule, it may be said that the search-find programs give the user the first cut at the literature by providing him with the standard indices that are usually applied in a simple search.

find title: superconduct
print: all

is the equivalent of what is usually called the KWIC index or key word in context. "Find author" and "find location" use TIP as author and address indices, respectively. "Find citation XXX" is the citation index technique. The contribution of TIP in this respect is that the information is available on-line and that we need not print more than what the user wishes to see. In the standard citation index, literally millions of words are printed and displayed to the user. The probability that any one piece of information is of interest to a given user is thus vanishingly small.

The ability to go from one index to another within a few seconds is a decisive factor that gives power to these indices. Thus, it is possible to string the commands into a sequence such as

find title: laser;

having found these, (all papers with laser in the title)

find all citations to them;

having found these,

find all other papers by the same authors;

having found these,

find all papers by others in the same laboratory, etc., etc.

We shall return to this in a later part of this report.⁴

2. Programs for Handling Citations. Recent interest in the citation literature as a significant sub-set of the total published literature made it desirable to develop a set of special programs that allow diverse manipulation of references and bibliographies carried in the published literature. We consider that a given issue of a journal is the carrier of two types of literature: the published literature, consisting of a set of titles whose text appears in the carrier issue, and a citation literature, consisting of a set of titles whose text is not in the carrier issue but is dispersed in time and place of publication.

⁴ "TIP System Applications", Appendix B.

The published literature may be described by a set of numbers giving the place (journal) and time (volume, page) of publication. This set of numbers remains fixed for all subsequent time. The citation literature, on the other hand, has no such fixed descriptors in time and place. Thus, given a number of published papers, we can state once and for all where and when they were published. Given the same group of papers, we cannot say exactly where and when they were or will be cited, nor can we expect any such statement to remain fixed in time. It is clear that special programs are necessary that will derive the cited literature from the published record and present it as a time-varying approximation.

Far from being a burdensome nuisance, this dynamic time-dependence of the citation literature is possibly its most interesting and powerful feature from the point of view of retrieval. The working literature of the active scientist is also dynamic and fluid with time. It is, therefore, possible that the working literature in a given problem situation is more easily derived from the citation literature than from the total published set of all papers.

Fig. 2 shows the relations between the various classes of literature. The TIP public library is a bounded region in journal-time space. If a_i is a given paper published at a given time, it defines three derivative groups of papers related to it by way of the citation mechanism. In Fig. 2, papers b, d, and f are cited by a_i . These may be retrieved from the TIP system by the "search-find" method. For example:

```
search all
find  $a_i$ 
print bibliography
```

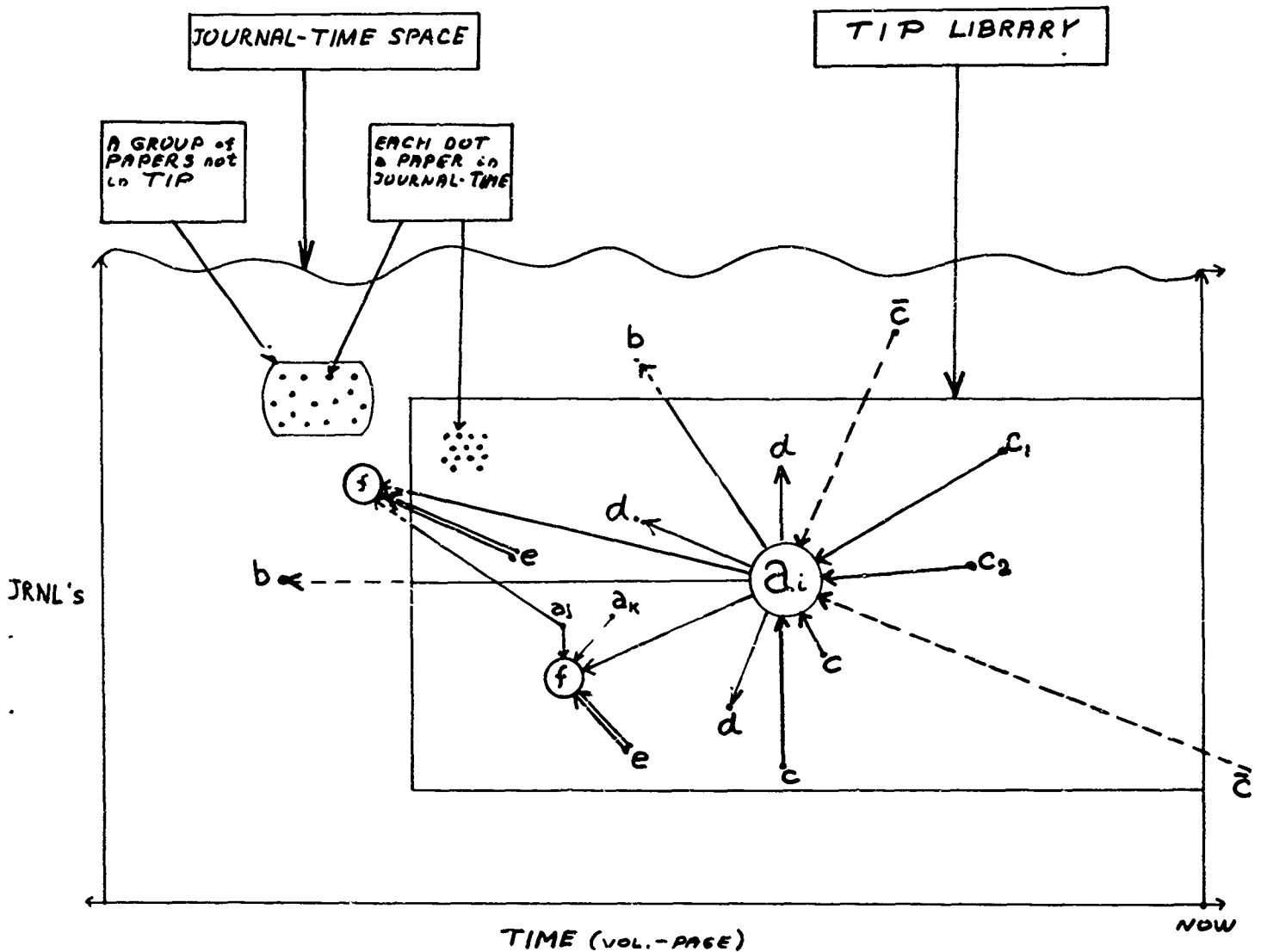
Various extensions of this capability will be discussed later in this report.

The papers in groups c and e are those that refer to or cite a_i . These, too, may be found easily in TIP. For example:

```
search all
find bibliography =  $a_i$ 
print all
```

In both cases, a_i would be identified in terms of jn., vol., page.

The above procedure applied to a large number of papers arranged in



a_l, a_s, a_k A GROUP OF PAPERS IN TIP = A
 \bar{c} CITATIONS TO A NOT IN TIP
 c CITATIONS TO A IN TIP
 b CITATIONS BY A NOT IN TIP
 d CITATIONS BY A IN TIP
 f MOST FREQUENT MEMBERS OF D AND B
 e CITATIONS TO F IN TIP
 c_1 and c_2 HAVE BIBLIOGRAPHIC COUPLING THROUGH a_i

Fig. 2

sequence constitutes a citation index. TIP contains several programs that create and manipulate citation index information.

Fig. 3 shows one possible output, in this case, a journal inventory. This is a count and a listing of each journal and the number of times that it is cited in a given body of literature. From Fig. 3 we learn that vol. 28 of Annals of Physics cites 336 articles. The first line tells us that The Physical Review (code number 1) was cited 90 times, while Physics Today (code number 98) was cited but once. Various manipulations may be performed on the citation index once it is produced and stored. In Fig. 4 we asked for the detailed listing of the 32 references to Nuovo Cimento. The indexing may be produced volume by volume, stored and then merged to produce an overall master citation index of the entire literature in the TIP library.

Perhaps the most powerful exploitation of the bibliographic record derives from a set of TIP programs that we call "share". In this process, we name an article or a set of articles and ask that others be found that share an element with it. One may ask for author, word, location, or bibliography share. It is the latter case that has proven to be most useful for retrieval and that we have most thoroughly investigated under the name of bibliographic coupling or "share b".^{5, 6} In Fig. 2, a_i and the paper marked "e" both contain a reference to "f". We, therefore, say that a_i and "e" are bibliographically coupled by virtue of paper "f". The following definitions will be useful.

- a. A single item of reference used by two papers is called one unit of coupling between them.
- b. The coupling strength between two papers is measured by the number of coupling units (n) between them.
- c. A number of papers constitute a related group G_A if each member of the group has at least one coupling unit to a given test paper P_0 .
- d. A number of papers constitute a related group G_B if each member of the

⁵Kessler, M. M., "Comparison of the Results of Bibliographic Coupling and Analytical Subject Indexing". American Documentation, 16, 3, pp. 223-233. (Appendix F).

⁶Kessler, M. M., "Bibliographic Coupling Extended in Time: Ten Case Histories". Inform. Stor. Retr., 1, pp. 169-187. (Appendix G).

INVENTORY OF CITATIONS IN ANNALS OF PHYSICS (384) v. 28

Index annphy 28

YOUR CITATION INDEX HAS BEEN CREATED
TOTAL NUMBER OF REFERENCES IS 336

count all

COMPLETE COUNT OF CITATIONS FOLLOWS

NREFS JCODE JOURNAL NAME

90	1	PHYS REV
22	2	BOOKS
1	4	ASTROPHYS J
2	5	ANN PHYS
5	12	J CHEM PHYS
3	13	J PHYS(USSR)
32	17	NUOVO CIMENTO
1	22	PROC CAMB PHIL SOC
8	23	PROC ROY SOC(LONDON)
24	27	PRIVATE COMM.UNPUBLISHED, TO BE PUBLISHED
2	28	PHIL MAG
4	29	PROGR OF THEORET PHYS (JAPAN)
4	30	REVS MOD PHYS
8	32	REPORTS, TECHNICAL MEMOS
1	33	THESES
8	34	Z PHYSIK
14	41	PHYS REV LETTERS
1	43	HELV PHYS ACTA
5	46	BULL AM PHYS SOC
3	49	PHYS LETTERS
1	51	Z NATUR FORSCH
2	56	ANN REV NUCLEAR SCI
2	74	AM J PHYS
1	76	REPTS PROGR IN PHYS
1	80	J PHYS SOC (JAPAN)
2	96	COMPT REND
1	98	PHYS TODAY
1	120	ZHUR FIZ KHIM (ZHFK, ZFKH)
2	182	PROC THEORET PHYS SUPPL
1	206	ZHUR EKSPTL I TEORET FIZ (JETP)
12	227	J MATH PHYS
22	384	ANN PHYS (N.Y.)

Fig. 3

LISTING OF CITATIONS TO NUOVO CIMENTO (17)
IN ANNALS OF PHYSICS (384) v. 28

print 17
CITATION INDEX FOR--
NUOVO CIMENTO

CITED ARTICLE			CITING ARTICLE		
J0017	V002	P0425	J0384	V028	P0466
J0017	V004	P0323	J0384	V028	P0034
J0017	V007	P0607	J0384	V028	P0060
J0017	V007	P0794	J0384	V028	P0400
J0017	V007	P0843	J0384	V028	P0400
J0017	V008	P0316	J0384	V028	P0466
J0017	V011	P0342	J0384	V028	P0466
J0017	V014	P0951	J0384	V028	P0320
J0017	V017	P0956	J0384	V028	P0435
J0017	V018	P0198	J0384	V028	P0034
J0017	V018	P0466	J0384	V028	P0466
J0017	V018	P0933	J0384	V028	P0034
J0017	V018	P0947	J0384	V028	P0320
J0017	V021	P0186	J0384	V028	P0034
J0017	V021	P0524	J0384	V028	P0466
J0017	V021	P1094	J0384	V028	P0018
J0017	V022	P0362	J0384	V028	P0034
J0017	V022	P0494	J0384	V028	P0034
J0017	V023	P0047	J0384	V028	P0225
J0017	V024	P0870	J0384	V028	P0466
J0017	V025	P0224	J0384	V028	P0466
J0017	V025	P1135	J0384	V028	P0466
J0017	V026	P1254	J0384	V028	P0034
J0017	V027	P0193	J0384	V028	P0034
J0017	V027	P0384	J0384	V028	P0034
J0017	V028	P1294	J0384	V028	P0034
J0017	V028	P1464	J0384	V028	P0034
J0017	V030	P0177	J0384	V028	P0320

Fig. 4

group has at least one coupling unit to at least one other member of the group.

- e. Definitions of G_A and G_B may be modified by assigning higher values of n .

Fig. 5 is an example of the "share b" program producing a G_A group. In this case, we searched The Physical Review, volumes 131 to 134 for all articles that share a reference with the test paper P_0 in Soviet Physics - JETP, volume 18, page 945. The title of P_0 was "The Mossbauer Effect on ^{119}Sn Nuclei in a Vanadium Matrix". For output we requested the titles, authors, and shared citations. The G_A thus produced contains nine articles related to P_0 with coupling strength " n " that varies between $n=1$ and $n=4$. Although the match is based entirely on bibliographic coupling, the remarkably close relation between the title of the test paper and those of nine papers retrieved is very clear. The relation goes beyond the mere sharing of words. Of the nine papers, three contain the word "Mossbauer" in the title, four relate to studies on metals and four deal with metallic crystal structure, lattice and lattice defects. All of these topics are implied in the title of the test papers.

Fig. 6 is a schematic representation of an actual G_B , although we shall not here examine the titles of each paper. In 6a, a group of 16 papers is formed on the basis of $n \geq 1$. In 6b, G_B is redefined for $n \geq 3$. The group breaks up into 3 smaller groups where the criterion of relatedness is much stronger. The validity of bibliographic coupling as a mechanism for retrieval has been studied and tested for several thousand cases. We shall return to it later in this report to show how it can be used as a building block or element in a complex search strategy. It may be of interest to point out at this time that the three modes of bibliographic utilization, namely the author's bibliography, the citation index and bibliographic coupling behave differently with regard to time. Given a paper P_0 , the author's bibliography is in the past, the citation index is yet to be generated, while the results of bibliographic coupling are time-symmetric, all, of course, with respect to P_0 . Fig. 7 illustrates the last point. A paper in volume 97 (1955) was chosen as P_0 . The search range

EXAMPLE OF BIBLIOGRAPHIC COUPLING

search phyrev v.131 to v.134
find shared bibliography spj etp v.18 p.945
output print title citations
go

PHYSICAL REVIEW

VOLUME: 131

PAGE: 0529

TEMPERATURE DEPENDENCE AND ANISOTROPY IN THE DEBYE-WALLER FACTOR
FOR WHITE TIN

J0730 V006 P0881

PAGE: 0535

LOCALIZED MODE DETECTION BY MEANS OF THE MOSSBAUER EFFECT

J0001 V129 P0028

PAGE: 1008

DYNAMICAL MOTION AND GAMMA-RAY CROSS SECTION OF AN IMPURITY NUCLEUS
IN A CRYSTAL. I. ISOLATED IMPURITIES IN GERMANIUM AND ALUMINUM

J0001 V126 P2059 J0001 V129 P0028

PAGE: 1500

PHONON SCATTERING BY LATTICE DEFECTS

J0001 V129 P0028

VOLUME: 132

VOLUME: 133

PAGE: 1062

NUCLEAR ZEEMAN EFFECT IN GOLD ATOMS DISSOLVED IN IRON, COBALT,
AND NICKEL

J0001 V123 P0816

PAGE: 1553

MOSSBAUER EFFECT FOR FE57 IN BERYLLIUM, COPPER, TUNGSTEN, AND
PLATINUM

J0669 V015 P0182 J0669 V017 P0195 J0001 V126 P2059
J0001 V129 P0028

VOLUME: 134

PAGE: 0716

LATTICE DYNAMICAL STUDIES USING ABSOLUTE MEASUREMENTS OF THE
LAMB-MOSSBAUER RECOIL-FREE FRACTIONS

J0001 V129 P0028 J0001 V126 P2059 J0669 V017 P0195

PAGE: 0965

LOCALIZED MODE IN AN ANHARMONIC CRYSTAL

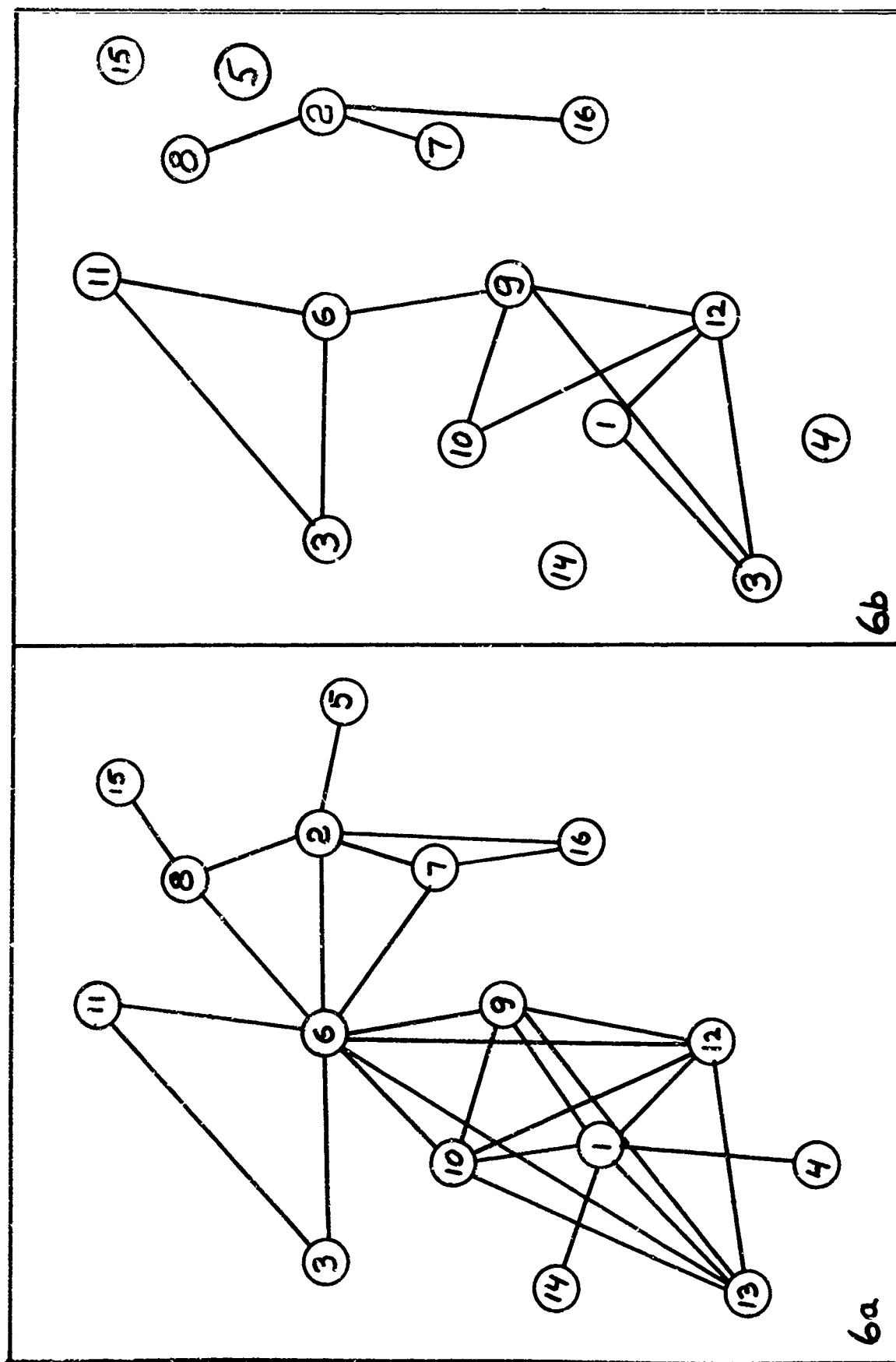
J0669 V015 P0182 J0669 V017 P0195 J0001 V129 P0028

PAGE: 1486

FREQUENCY SPECTRA OF BODY-CENTERED CUBIC LATTICES

J0030 V030 P0250 J0001 V109 P1046

Fig. 5



DIAGRAMATIC REPRESENTATION OF G_B

Fig. 6

Number of papers

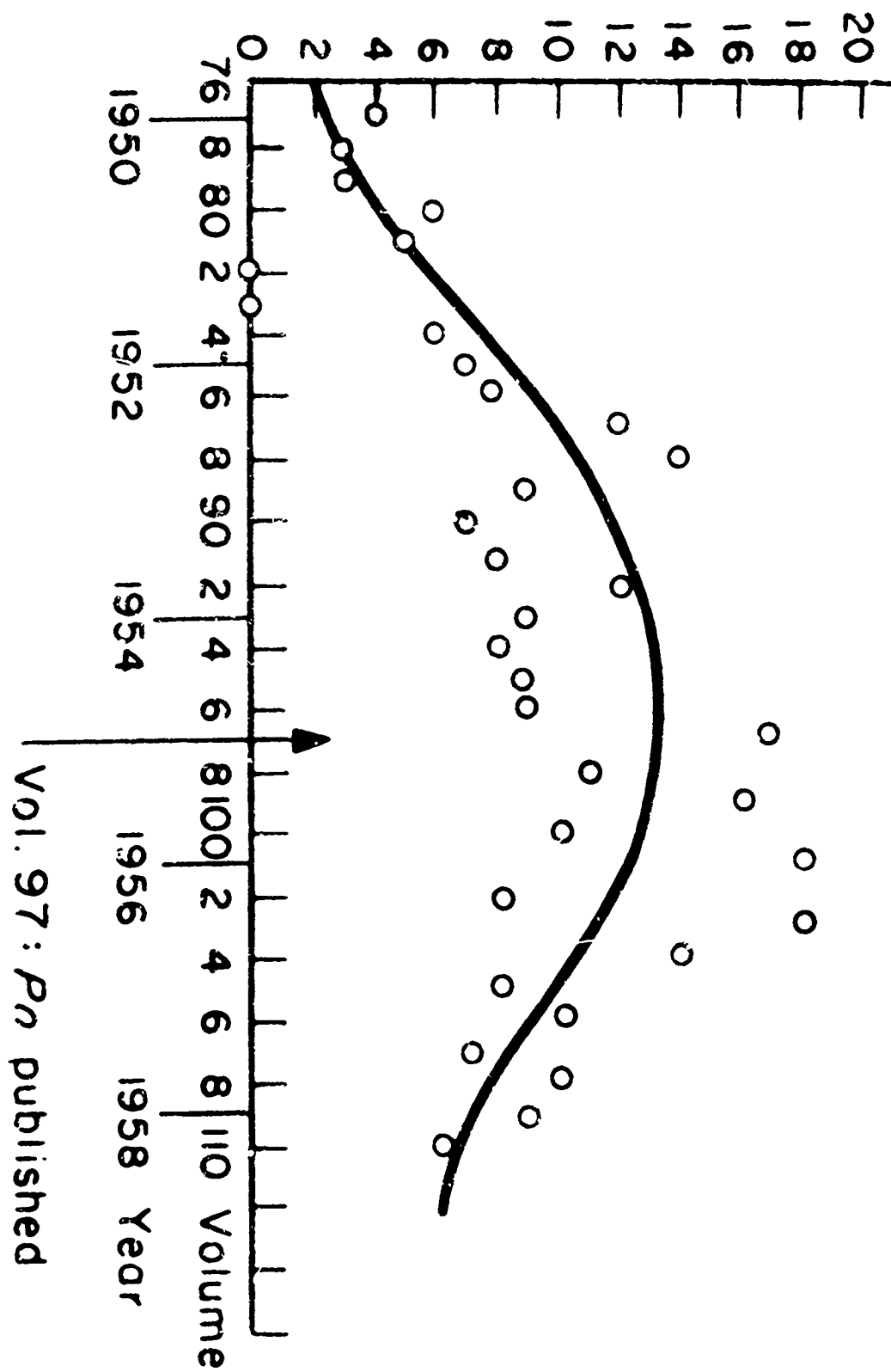


Fig. 7

extended over an eight-year period from 1950 to 1959. 322 papers were retrieved into this group. Fig. 7 shows the distribution of these papers on either side of P_0 . The retrieved papers are fairly symmetrically distributed on either side of the volume containing P_0 . The search span of about four or five years to either side of the test paper was not arbitrary. It was adjusted automatically by the search method and reflects the so-called half-life of citations in the published literature. This provides a self-purging effect on papers in G_A and G_B , insuring that the list is eventually closed and does not continue to grow without limit.

3. Output: Print, Store, Save. The TIP search and retrieval programs described above will result in lists of papers varying in size from a few to several hundred. If the list is small, the on-line user will usually wish to see the results printed immediately on the console. In this case, the command is "output print" followed by the information desired. Thus, in Fig. 1 we asked for "output print: title, author, location", etc. Any combination of the recorded elements of the document may be asked for. The print-out is in a standard, prearranged format and will include at the end a statement on how many articles were searched, how many were found, and the computer time used in the operation.

It may happen that the list of papers produced by a search is expected to be too large or that for some other reason an immediate print-out at the console is not desired. In that case, we have the option of storing the output for future print or display. The file thus stored is given a name that can later be used to identify it. Thus, following a search procedure we may type:

```
output store all  
name stored file list A.
```

The results of the search will be stored in the computer under the title "list A". This list is available only to the user who compiled it and will remain in his private files. Several forms of output are now at the user's disposal. The command "request print list A" will initiate a print-out of the material on the high-speed printer in the computer room at a time, usually late at night, when the computer is not in demand. The print-out may be picked up the following morning.

request punch list A
will produce the information on punched cards which, if desired, may be converted to tape. When the information is no longer wanted, the originator may erase it with the command

delete list A

The above "store" command prepares the information for the last output stage. In this format the list is decoded and expanded into a format suitable for output but not for further TIP manipulation.

It frequently happens that a list is produced by a user who wishes to perform additional TIP manipulations on it. This is frequently the case when a search produces a large sub-group of the total library which is then manipulated for finer-grade retrieval. For example,

search all
find title laser

will search the entire TIP library of 30,000 journals and produce perhaps 400 papers, each with the word laser in the title. We may then want to further search this list for some specific group of papers, say all those that share bibliography with a given paper. To save a list for subsequent TIP manipulation we use the command

output save
name saved file laser list.

The user may then apply all the TIP programs to the saved list. For example:

search: laser list
find title: hydrogen
author: Smith
share b: p_1 p_2
output print all

will examine the list of papers with laser in the titles and remove all those that contain also the word "hydrogen" or are by author Smith or share one or more references with papers P_1 and P_2 . Saved files are available only to the user of origin and will remain in his private file until deleted by him or transferred

to the public files in which case they become an integral part of the TIP public library available to all TIP users.

4. List Handling Programs. A group of programs has been developed by TIP that facilitates the handling and arrangement of large lists of documents. These programs may be applied to files in the TIP library or to saved files that result from a previous manipulation of TIP material or, indeed, to any file that exists in TIP format regardless of its origin or public status.

Tally: This program will arrange elements of a paper in alphabetic order and give the count of each element. "Tally title", for example, will yield an alphabetic list of words in titles and count the frequency of each word. "Tally author" will give an alphabetic list of authors and their frequency of occurrence. Similarly for "tally location" and "tally citations". The tally listing is ordinarily arranged in alphabetic sequence with the frequency shown as it occurs. A program called "freq." will rearrange a tally listing by descending or ascending order of frequency. The use of this facility will be discussed later as an element in the construction of a complex search strategy.

Sort: The "tally" program gives statistical information only; it does not produce an index. It may at times be of interest to know not only how many times a word exists in a file of titles but exactly what papers contain the given word. The program "sort" produces this information. "Sort title" will list each word and the citations to the papers that contain it. "Sort author", "sort location", and "sort citations" will produce, respectively, an author index, a location index, and a citation index. The output in each case may be requested to be only the citations (jn., v., p.) or any combination of recorded elements.

Merge: This program will combine or merge two lists according to some criterion. For example, an author index as described above may be created for each volume of a journal. When several volumes are so indexed, the lists may be merged to form one master index. In general, several lists of papers may be formed according to different but related criteria. It may then be of interest to combine the lists and eliminate duplications. The "merge" programs are generally useful

in up -dating, editing, adding and deleting listed information.

Permute: This program will arrange a list of papers into related groups according to the criterion of bibliographic coupling. Each paper in turn is regarded as the test paper P_0 against the remaining members of the group.

This is a most powerful procedure when applied to a list with some a priori probability of relatedness. For example,

```
search all
find title lasers
save all
name saved list laser titles
```

will produce a master list of papers all having the word "laser" in the title.

```
search laser titles
permute
```

will break up the master list into a number of sub-groups G_A such that interest in any one paper in the group implies a high probability that the rest of the group should be looked at. This process is equivalent to an indexing where the index element, instead of being a word or an author or a citation, is a paper or a group of papers.

5. Read or See Also. It is frequently convenient to group a set of commands and instruction under a new name and then be able to call for it as one entity instead of having to repeat each component separately. For example, the TIP public library consists of a three-year collection of 38 journals. The journals, although all pertaining to physics, nevertheless cover a very large range of subjects. A given user may know from his own experience that the bulk of his material is likely to be found in smaller groups of journals. Furthermore, he may want to confine his operations to a certain time span. He may form a saved file of the sub-library of interest to him and name it in the following manner.

```
search: physical review vol. 140 to 146
      physical review letters vol. x to y
      journal of applied physics vol. x to y
      physics letters vol. x to y
      nuovo cimento vol. x to y
save all
name saved file my file
```

Every time a user wants to search the above literature he need not re-type the search request. The instruction may be activated by typing "read my file" followed by the appropriate search and output command.

Indeed, the "read" command may consist of an entire package of search, find, and output requests, as in the following:

```
search: all new
find title: w1, w2, w3...
          authors: a1, a2, a3...
          citations: p1, p2, p3...
          share b: p1, p2, p3...
output print all and save
name saved file my file
```

Periodically, say once a week, the user may type the single line
read my file.

The computer will then examine the "new literature" and retrieve all papers that contain the given words, authors, citations, etc. as enumerated in the user's read file. The read file is always under the user's control who may edit and up-date it at any time by adding, deleting or expanding his list of instructions. This will be recognized as an example of selective dissemination of information where the criteria of selection are entirely under the user's on-line control.

6. User-Originated Input. So far, the user has operated on files from the TIP Public Library or on saved files created in some way from the TIP libraries. It may at times be desirable for a user to create a file of his own choosing or to add to an existing file of TIP origin. Thus, in a given specialized field a user may want to add other documents to the TIP journals or perhaps put his private collection of reprints into TIP format.

The MAC-TIP system provides on-line input facilities for storing and editing information on the discs. If this information is stored in proper format, all the TIP search and retrieval programs may be applied. This aspect of TIP will be treated in more detail in a subsequent section.

7. Expanded Field Possibilities. The programs and procedures developed by TIP are quite general and not at all sensitive to the nature of the information

that is being searched or processed. Thus, it is not confined to the technical and scientific literature nor, indeed, to literature as a class. Any information that is of a list or serial nature consisting of identifiable items and properties pertaining to them may be stored and searched with the TIP programs. For example, instead of scientific papers with authors, titles, etc. we may want to create and search a personnel file where the identifiable item is the name of the person and the pertinent information consists of his address, status, education, salary, etc., etc. or if a student, his academic standing, courses, fellowships, etc., etc. In this case, obviously the field designations title, author, etc. do not apply nor is it always convenient to squeeze all the information into the five fields used by TIP for literature storage and retrieval. For this purpose, a generalized TIP-like package is available that allows the user to name his items in any way that is appropriate and to assign and name as many fields as the information requires. All the TIP programs are applicable to these files.

8. Monitor. The personnel of Project TIP may from time to time wish to monitor the entire system or portions thereof in order to ascertain how and by whom certain programs are being used, to observe the system response and detect difficulties either in the system or in the way that it is being used by the outside public. For this purpose we have written a monitor program that may be turned on and off as desired. When the monitor is on, a periodic printout is made available to the TIP management that records the various engagements that the system had been subjected to in the previous day or week. It provides a user record without in any way engaging the user with questions, interviews, or other disturbing influences.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL INFORMATION PROGRAM *
AND PROJECT MAC **

TIP USER'S MANUAL

A guide for on-line search and retrieval
of the current literature in physics.

M. M. Kessler

FIRST EDITION	December, 1965
SECOND PRINTING	December, 1966
REVISED	October, 1967

*Supported by a grant from the National Science Foundation

**Sponsored by the Advanced Research Projects Agency,
Department of Defense, under Office of Naval Research

TIP USER'S MANUAL

Table of Contents

Preface	1
Introduction	3
Library	5
Search	7
Find	9
Share B	15
Output Print	20
Output Save	20
Output Store	23

PREFACE

The Technical Information Program (TIP) provides a means for on-line search and retrieval of a large body of the physics journal literature.

The library in store on the MAC-CTSS discs consists of the following journals:

ANNALS OF PHYSICS
APPLIED PHYSICS LETTERS
CANADIAN JOURNAL OF PHYSICS
CHINESE JOURNAL OF PHYSICS
DISCUSSIONS OF THE FARADAY SOCIETY
HELVETICA PHYSICA ACTA
INDIAN JOURNAL OF PHYSICS
JAPANESE JOURNAL OF APPLIED PHYSICS
JETP LETTERS
JOURNAL OF APPLIED PHYSICS
JOURNAL OF CHEMICAL PHYSICS
JOURNAL OF CHEMICAL PHYSICS (SUPPLEMENT)
JOURNAL OF MATHEMATICAL PHYSICS
JOURNAL OF THE OPTICAL SOCIETY OF AMERICA
JOURNAL OF THE PHYSICAL SOCIETY OF JAPAN
MOLECULAR PHYSICS
NUCLEAR PHYSICS (SERIES A)
NUCLEAR PHYSICS (SERIES B)
NUOVO CIMENTO (SERIES A)
NUOVO CIMENTO (SERIES B)
NUOVO CIMENTO (SUPPLEMENTO)
PHILOSOPHICAL MAGAZINE
PHYSICA
THE PHYSICAL REVIEW
PHYSICAL REVIEW LETTERS
PHYSICS LETTERS (SERIES A)
PHYSICS LETTERS (SERIES B)
PHYSICS OF FLUIDS
PROCEEDINGS OF THE PHYSICAL SOCIETY
PROCEEDINGS OF THE ROYAL SOCIETY
PROGRESS OF THEORETICAL PHYSICS
PROGRESS OF THEORETICAL PHYSICS (SUPPLEMENT)
SOVIET JOURNAL OF NUCLEAR PHYSICS
SOVIET PHYSICS - JETP
SOVIET PHYSICS - SEMICONDUCTORS
SOVIET PHYSICS - SOLID STATE
SOVIET PHYSICS - TECHNICAL PHYSICS
TRANSACTIONS OF THE FARADAY SOCIETY

The holdings start with the first issue of 1965 and are updated weekly.

For each article in each of the above journals the disc contains the following information:

- i - identification (journal, volume, page)
- t - title
- a - author(s)
- l - author's institutional connection
- c - citations and bibliography

About 30,000 articles are now in disc storage.

The literature is arranged by journal and volume. The search language allows the location and retrieval of a set of papers that contain a given item or satisfy a set of conditions.

This manual is written for the MAC user who is interested in literature retrieval. We assume that the user has access to the MAC system and understands how to LOGIN, QUIT, and LOGOUT.

The information in this manual is presented largely as a series of examples of actual search problems. The user is advised to repeat each example as it is given thus learning by doing. Additional information may be found in section AH.2.16 of the CTSS manual.

TIP in its present form is experimental and subject to change. As new capabilities are developed, they will be incorporated into the system and announced. An important guide to the improvement of TIP will be the feedback from the users. We request that all users who discover faults in the data, difficulties with the programs or who have questions and suggestions regarding TIP, please write to Project TIP, M.I.T. Libraries, Room 14S-310.

INTRODUCTION

After "login" the user types "tip" in order to engage the literature search facility of the computer.

As a general rule, the TIP user must give three instructions in order to elicit a response. The three instructions are:

- SEARCH - defines the literature range to be examined.
- FIND - defines the object of the search, the item to be found, or the condition to be satisfied.
- OUTPUT - defines what the user wishes to have associated with the item found and how it is to be displayed to him.

Example 1. Illustrates a user who wants to examine volumes 26 to 28 of the ANNALS OF PHYSICS and find all papers that have the word "pion" in the title. For output, the user is requesting an immediate print-out of the identifications, titles, and authors.

tip
W 1019.5
TYPE YOUR REQUESTS.

search annals of physics v.26 to v.28
find title pion
output print title author identification and location
go

ANNALS OF PHYSICS
VOLUME 26
VOLUME 27
J384 V027 P0079
DEUTERON PHOTODISINTEGRATION AND N-P CAPTURE BELOW PION
PRODUCTION THRESHOLD
PARTOVI F.
CAMBRIDGE, MASSACHUSETTS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LABORATORY FOR NUCLEAR SCIENCE AND PHYSICS DEPARTMENT

VOLUME 28
J384 V028 P0034
ANALYSIS OF THE PHOTOPRODUCTION OF POSITIVE PIONS
HOHLER G.
SCHMIDT W.
GERMANY
TECHNISCHE HOCHSCHULE KARLSRUHE
INSTITUT THEORETISCHE KERNPHYSIK

J384 V028 P0433
MAJORIZATION OF FEYNMAN DIAGRAMS INVOLVING PIONS AND NUCLEONS
BOYLING J. B.
CAMBRIDGE, ENGLAND
CAVENDISH LABORATORY
DEPT. OF APPLIED MATHEMATICS AND THEORETICAL PHYSICS

Notes on Example 1:

1. No articles with "pion" were found in volume 26, one was found in volume 27, page 79 and two in volume 28, page 34 and page 435. J384 is the numeric code for ANNALS OF PHYSICS.
2. By asking for "pion" we also found titles with "pions". In general, everything beginning with "pion" will be found.
3. The asterisk is interpreted as an explicit blank. Therefore, "find title pion*" will find only pion and not pions or pionization.
4. "Find title +pions" will find pion or pions with any prefix such as "hyperpions".
5. "Find title thin film" will find all titles containing the words "thin" and "film" in any order. To insist on the exact order, type "find title thin*film". The * insists that there be nothing but a blank between the indicated words.
6. Be sure you understand the following find options:

find title	conduct
find title	conduct*
find title	+conduct
find title	+conduct*
find title	high temperature
find title	high*temperature

You may now practice TIP on ANNALS OF PHYSICS. Volume 26 to 34 are on the disc.

A general note:

If a typing error is made, use the single quote (") to erase the letter immediately preceding, or as many quotes as are necessary to erase preceding letters and continue typing. Use the question mark (?) to erase the preceding line and continue typing.

find title conducp"t - will erase the p and substitute the t
find title condulp""ct - will erase both l and p and substitute ct
find tiple conduct?find title conduct - will erase the entire
line preceding the (?) and accept the line following the (?)

LIBRARY

Example 2 illustrates a user asking for the inventory of journals and volumes currently in TIP. To obtain the list, type "library".

library

THE DATE IS 12/13/65
THE VOLUMES AVAILABLE TODAY ARE...

ANNALS OF PHYSICS

J384 - ANNPHY - ANN PHYS
V 26 - 34

APPLIED PHYSICS LETTERS

J646 - APPLLET - APPL PHYS LETTERS
V 4 - 7

CANADIAN JOURNAL OF PHYSICS

J55 - PHYCAN - CAN J PHYS
V 41 - 43

HELVETICA PHYSICA ACTA

J43 - PHYHEL - HELV PHYS ACTA
V 36 - 38

INDIAN JOURNAL OF PHYSICS

J164 - INDJPH - IND J PHYS
V 38 - 39

JETP LETTERS

J821 - JETLET - JETP LETTERS
V 1

JAPANESE JOURNAL OF APPLIED PHYSICS

J612 - PHAPJA - JAPAN J APPL PHYS
V 2 - 4

JOURNAL OF APPLIED PHYSICS

J11 - PHYAPP - J APPL PHYS
V 34 - 36

JOURNAL OF CHEMICAL PHYSICS

J12 - JCHEPH - J CHEM PHYS
V 38 - 43

JOURNAL OF MATHEMATICAL PHYSICS

J227 - MATHPH - J MATH PHYS
V 6

JOURNAL OF THE PHYSICAL SOCIETY OF JAPAN

J80 - PHYSOJ - J PHYS SOC JAPAN
V 18 - 20

continued on next page

Example 2 continued from page 5.

IL NUOVO CIMENTO

J17 - NUOCIM - NUOVO CIMENTO
V 27 - 38

NUCLEAR PHYSICS

J682 - NUCPHY - NUC PHYS
V 50 - 69

PHYSICA

J21 - HYSICA - PHYSICA
V 29 - 31

PHYSICS LETTERS

J49 - PHYLET - PHYS LETTERS
V 3 - 17

THE PHYSICS OF FLUIDS

J799 - PHYFLU - PHYS FLUIDS
V 6 - 8

PHYSICAL REVIEW

J1 - PHYREV - PHYS REV
V 129 - 139

PHYSICAL REVIEW LETTERS

J41 - PHYRET - PHYS REV LETTERS
V 10 - 15

PHYSICAL REVIEW, SERIES B

J199 - PHYREB - PHYS REV B
V 133 - 139

PROCEEDINGS OF THE PHYSICAL SOCIETY (LONDON)

J3 - PHYPRO - PROC PHYS SOC
V 81 - 85

PROGRESS OF THEORETICAL PHYSICS (KYOTO)

J29 - PROPHJ - PROGR THEORET PHYS
V 29 - 34

SOVIET PHYSICS-JETP

J669 - SPJETP - SOVIET PHYS JETP
V 17 - 21

SOVIET PHYSICS-SOLID STATE

J310 - SPSOLS - SOVIET PHYS SOLID STATE
V 5 - 6

SOVIET PHYSICS-TECHNICAL PHYSICS

J790 - SPTPHY - SOVIET PHYS TECH PHYS
V 8 - 10

END OF LIBRARY

end of Example 2.

Notes on Example 2:

1. The library list shows the four ways in which a user may name a journal for TIP.
 - a. The full name of the journal (ANNALS OF PHYSICS).
 - b. A numeric code, always preceded by the letter J (J384).
 - c. A six letter code (ANNPHY).
 - d. The abbreviation of the journal as recommended by Chemical Abstracts (ANN PHYS).
2. The user is urged to settle on one method of naming the journal and to get into the habit of using it.
3. The printed list will be updated automatically as new material is added to the library.
4. The reasons for choosing this particular list of journals for TIP and the criteria to be used for the admission of new journals are discussed in an article by M. M. Kessler, Physics Today, March 1965, page 26.
5. Limitations of presently available MAC disc storage forced us to stop the TIP literature at 1963. For many of the 25 journals, machine useable data is available back to 1960 and in the case of the Physical Review back to 1950. As more storage space becomes available, we expect to maintain the library on the basis of at least five years.

SEARCH

A search range is specified by naming the journal and volume span to be examined. The journals may be named by any of the four methods given above. Several options are available for stating the search command.

- a. Search PHYREV 133
- b. S Phyrev 135 to 135
- c. S Phyrev (this command means search all of the available volumes of the named journal)
- d. S all (will search the entire TIP library)
- e. S Ja, Jb, Jc Vm to Vn, Jd Vp (will search all of the journals a and b, volumes m to n of journal c and volume p of journal d)

Example 3 illustrates a multiple search command.

s annals of physics 28, applied physics letters 4 to 6, j55 41
s indian journal of physics, phyrev 96, economic review
f t nuclear
o n i t
?

ANNALS OF PHYSICS
VOLUME 28
J384 V028 P0018
ELECTRODYNAMIC PROCESSES WITH NUCLEAR TARGETS

APPLIED PHYSICS LETTERS
VOLUME 4
VOLUME 5
J646 V005 P0116
3HE NUCLEAR ZEEHAN MASER

VOLUME 6

CANADIAN JOURNAL OF PHYSICS
VOLUME 41
J055 V041 P0762
NUCLEAR CHARGE DISPERSION IN THE HIGH-ENERGY FISSION OF
URANIUM

J055 V041 P1580
NUCLEAR SPIN RELAXATION IN GASES AND LIQUIDS II. MOLECULAR
HYDROGEN

J055 V041 P1629
NUCLEAR MAGNETIC RESONANCE STUDY OF FERROELECTRIC LITHIUM
HYDRAZINIUM SULPHATE, $Li(N_2H_5)SO_4$

INDIAN JOURNAL OF PHYSICS
VOLUME 38
J164 V038 P0326
ON THE DETERMINATION OF DISTORTION IN NUCLEAR EMULSIONS

J164 V038 P0499
LOW ENERGY NUCLEAR BURSTS OF COSMIC RAYS AT SEA LEVEL

VOLUME 39

PHYSICAL REVIEW
FILE PHYREVV00096 IS NOT IN THE LIBRARY.

SEARCH COMPLETED, 538 ARTICLES.
9.28 SECONDS, 57.9 ARTICLES/SEC.
7 ARTICLES FOUND.

Notes on Example 3:

1. The command words "search, find, output, print, go" and the descriptions "identification, title, author, location, citation" may all be abbreviated to their respective first letters. The word "volume" may be omitted in the search command.
2. When nothing is found in a given volume of a journal, the journal name and volume are nevertheless printed in the output to reassure the reader that the search was made.
3. If the name of a journal is misspelled or if a journal or volume is asked for that is not currently in TIP, the proper statement will be made by the computer. In our example, Economic Review and volume 96 of Physical Review are not in TIP.
4. We will later describe how a user may create and name private files of his own. These can subsequently be searched with TIP in the same manner as the common library files.
5. The end of a search schedule is indicated by the statement "search completed" and some pertinent statistical quantities.
6. If for some reason the user wishes to stop the print out before "search completed", execute QUIT. The user now has two options.
 - a. To continue the process from the point of QUIT, type "start".
 - b. To erase the process to the point of QUIT, type "tip" and give new instruction.

FIND

The items of information for each article stand in five fields:

- i - identification: journal, volume, page
- t - title
- a - author(s): not more than ten
- l - location: institution where work was done
- c - citations: journal, volume, page

Example 4. A typical article in TIP storage.

```
search phyrev 135
find identification j1 135 960
output print all
go
```

```
PHYSICAL REVIEW
VOLUME 135
J001 V135 P0960
VIBRATIONAL AND CENTRIFUGAL EFFECTS ON THE MAGNETIC
SUSCEPTIBILITY AND ROTATIONAL MAGNETIC MOMENT OF THE HYDROGEN
MOLECULE
```

```
CHAN SUNNEY I.
PASADENA, CALIFORNIA
CALIFORNIA INSTITUTE OF TECHNOLOGY
GATES AND CRELLIN LABORATORIES OF CHEMISTRY
IKENBERRY DENNIS
DAS T. P.
RIVERSIDE, CALIFORNIA
UNIVERSITY OF CALIFORNIA
DEPARTMENT OF PHYSICS
```

J001 V034 P0057	J001 V041 P0713	J001 V041 P0721
J001 V058 P0310	J001 V078 P0711	J001 V080 P0476
J001 V085 P0937	J001 V087 P1075	J001 V094 P0350
J001 V094 P0893	J001 V103 P1254	J001 V111 P0203
J001 V112 P1929	J001 V115 P0897	J001 V126 P0146
J002 V000 P0000	J002 V000 P0000	J003 V065 P0178
J012 V009 P0061	J012 V019 P1030	J012 V020 P0527
J012 V021 P2070	J012 V023 P1131	J012 V032 P0105
J012 V035 P1065	J012 V035 P1967	J012 V037 P0214
J012 V037 P1527	J012 V038 P1263	J027 V000 P0000
J027 V000 P0000	J030 V032 P0231	J030 V035 P0130
J046 V006 P0019	J055 V035 P0730	J160 V004 P0061
J311 V003 P0017	J311 V010 P0278	

Notes on Example 4:

1. The titles, authors and location are alphabetized. The identifications and citations are stored in numeric code. J001 V034 P0057 refers to a paper in volume 34 page 57 of the journal whose code number is 1 (Physical Review).
2. For the journals included in TIP, the J code numbers are given in the "library" list.
3. If a citation is to a paper within TIP's range of journal-volumes, it may be retrieved as illustrated in this example.

A simple find command consists of naming the field and defining the item within it.

```
f i J1 135 960
f t nuclear resonance (see discussion of + and * on page 4)
f a Smith
f a Smith J.P.
f l Massachusetts Institute of Technology
f l Lincoln Laboratory
f l India
f c J1 34 57
f c Phyrev 34 57
```

When find commands are arranged in consecutive lines as above, the word "or" is implied between each line. The "or" relation may also be stated explicitly as follows:

```
f t symmetric group or author Cooper R.K. or a Cooper Richard
or l Tucson, Arizona.
```

Example 5 illustrates the "or" relation.

```
tip
W 1700.6
TYPE YOUR REQUESTS.
```

```
s annals of physics 26
f t symetric group or author cooper r. k. or author richard cooper
or location tucson, arizona
o p i t a l
go
```

```
ANNALS OF PHYSICS
VOLUME 26
J384 V026 P0222
APPLICATION OF THE THEORY OF THE SYMMETRIC GROUP TO THE
SEVERAL-NUCLEON PROBLEM
MAHMOUD HORMOZ
COOPER RICHARD K.
TUCSON, ARIZONA
UNIVERSITY OF ARIZONA
PHYSICS DEPARTMENT
```

```
SEARCH COMPLETED, 22 ARTICLES.
.99 SECONDS, 22.2 ARTICLES/SEC.
1 ARTICLES FOUND.
```

When find commands are arranged linearly on one line, the word "and" is implied between the field elements. Thus:

1 t binary collision author Nelkin

is a request for a paper that has the word binary and the word collision in the title and also has Nelkin as one of the authors. The "and" relation may also be stated explicitly. Thus:

f t binary collision author Nelkin Mark
and location Cornell University

It is often useful to locate a set of papers that refer to a list of known papers. For this purpose, we use the command find citation J V P.

Example 6 shows a citation search.

s phyrev 135
f c j1 32 110
f c j1 126 912
o p i t linkage
g

PHYSICAL REVIEW
VOLUME 135
J001 V135 P0001
CURRENT FLUCTUATIONS IN A SUPERCONDUCTING CIRCUIT CARRYING A
CIRCULATING CURRENT
CITATION MATCH
J001 V032 P0110
CITATION MATCH
J001 V126 P0912

J001 V135 P1202
MICROWAVE SURFACE IMPEDANCE OF SUPERCONDUCTORS OF THE SECOND
KIND
CITATION MATCH
J001 V126 P0912

SEARCH COMPLETED, 266 ARTICLES.
2.79 SECONDS, 95.3 ARTICLES/SEC.
2 ARTICLES FOUND.

s phyrev 134
g

PHYSICAL REVIEW
VOLUME 134
J001 V134 P0001
MICROWAVE MAGNETOIMPEDANCE OF SUPERCONDUCTING TIN
CITATION MATCH
J001 V126 P0912

SEARCH COMPLETED, 222 ARTICLES.
2.39 SECONDS, 92.8 ARTICLES/SEC.
1 ARTICLES FOUND.

Notes on Example 7.

1. The "and" "or" "not" relations may be combined in one find command. When that is done, "and" always takes precedence to "or"; "not" is always assumed to be "and not". Find t a and b or c not d will be interpreted as (a and b) or (c not d).
2. Again, note how in the second part of example 7 the find command was changed leaving the search and output commands intact.

Example 8 illustrates one method of getting the table of contents of a physics journal.

```
s indian journal of physics 38
find author not iyndon*johnson
output p i t a
g
```

INDIAN JOURNAL OF PHYSICS

VOLUME 38

J164 V038 P0001

THERMAL CONDUCTIVITY OF THE SLOWLY REACTING SYSTEM $2H_1 H_2 + I_2$

SRIVASTAVA B. N.

CHADRABORTI P. K.

J164 V038 P0007

DOUBLE EMISSION OF HEAVY FRAGMENTS (Z EQUAL TO OR GREATER THAN 3) IN THE DISINTEGRATION OF EMULSION NUCLEI

DEKA G. C.

PATHAK K. M.

J164 V038 P0010

ELECTRICAL CONDUCTIVITY OF SINGLE CRYSTALS OF ILMENITE

MUKERJEE A. K.

J164 V038 P0019

A CONVENIENT AND DIRECT METHOD OF FINDING THE PRINCIPAL IONIC SUSCEPTIVITIES OF TRICLINIC CRYSTALS

GHOSH U. S.

MITRA S.

J164 V038 P0028

X-RAY METHOD OF DETERMINING THE AMPLITUDE FACTORS OF ORGANIC LIQUIDS

RATHO T.

TORASIA S.

MOHANTY J. C.

J164 V038 P0031

A NEW APPROACH TO THE SOLUTION OF SWITCHING FUNCTIONS HAVING CYCLIC PRIME IMPLICANT TABLES

CHOUDHURY A. K.

DAS SUNIL RANJAN

BASU M. S.

J164 V038 P0061

AN ANALYSIS OF THE J-PHENOMENON IN X-RAYS

PAL HIRENDRA KUMAR

SHARE B

A very useful retrieval method is to name one or more papers of known value to the user and ask that the literature be searched for other papers that share one or more citations with the given papers. This method, called bibliographic coupling, has been used and tested many hundreds of times and has been found most effective. A full description of this search tool may be found in Information Storage and Retrieval, Volume 1, pages 169-187, 1963 and American Documentation, Volume 16, page 223, July 1965.

Assume that a given user has two papers, P1 and P2, that are representative of his needs and interests. The two papers are j1 v1 p1 and j2 v2 p2. To search the literature for papers related to P1 and P2 through bibliographic coupling type:

s range of literature
f share b j1 v1 p1 and share b j2 v2 p2
o p t i linkage

The linkage in this case will be the shared reference.

Example 9a, b, c, d, e and f illustrates various aspects of the share b program.

We start with 9a, a search of physical review volume 135 for all titles that have the word laser in them. Two papers were found, P1 = j1 v 135 p316 and P2 = j1 v135 p578. Let us assume that P1 is of interest to a particular user.

In 9b, we search j1 v135 for papers that share one or more references with P1. Eight papers are thus retrieved. Note the close relation between the titles of these papers and the title of P1.

Example 9c illustrates the process of share b with P1 or share b with P2. Note that the two groups of shared papers are easily separated because the linkage information is given for each paper found.

Example 9d illustrates the process of share b with P1 and share b with P2. This is a much stronger condition and many of the papers found in 9c are eliminated.

Example 9e shows the process of find share b with a paper, but only those that do not contain a certain word in the title. As a general rule, the share b request may be mixed with other find requests in "or", "and", "not" relation.

Example 9f illustrates the process of finding share b when the test paper is not in the journal range being searched.

Example 9a and b. Use of share b.
s phyrev 135
f t laser
o p i t
g

PHYSICAL REVIEW
VOLUME 135
J001 V135 P0316
DOUBLE-RESONANCE PHENOMENA IN THE GASEOUS LASER

J001 V135 P0578
LASER SPECTROSCOPY OF A PULSED MERCURY-HELIUM DISCHARGE

f share b phyrev 135 316 but not i j1 135 316
o p t i linkage
g

PHYSICAL REVIEW
VOLUME 135
J001 V135 P0247
ELECTRON SPIN DOUBLE RESONANCE STUDIES OF F CENTERS IN KCL. I
SHARED LINKAGE TO PHYREV V135 P0316
J001 V070 P0460 J030 V026 P0167

J001 V135 P0543
FREQUENCY PUSHING AND FREQUENCY PULLING IN A HE-NE GAS OPTICAL
MASER
SHARED LINKAGE TO PHYREV V135 P0316
J001 V126 P0580

J001 V135 P0568
STEADY-STATE STIMULATED EMISSION IN RESONANT AND NONRESONANT
CAVITIES
SHARED LINKAGE TO PHYREV V135 P0316
J001 V126 P0580

J001 V135 P0582
SPIN RELAXATION OF OPTICALLY PUMPED CESIUM
SHARED LINKAGE TO PHYREV V135 P0316
J001 V076 P0833 J001 V107 P1559 J096 V229 P1213

J001 V135 P0591
STUDY OF SPIN-EXCHANGE COLLISIONS IN VAPORS OF RB85, RB87, AND
CS133 BY PARAMAGNETIC RESONANCE
SHARED LINKAGE TO PHYREV V135 P0316
J001 V070 P0460

J001 V135 P1068
SPIN-LATTICE RELAXATION IN FREE-RADICAL COMPLEXES
SHARED LINKAGE TO PHYREV V135 P0316
J001 V070 P0460

J001 V135 P1218
RELATION OF CROSSING AND ANTICROSSING EFFECTS IN OPTICAL
RESONANCE FLUORESCENCE
SHARED LINKAGE TO PHYREV V135 P0316
J041 V003 P0420

J001 V135 P1498
FORCED TWO-LEVEL OSCILLATOR
SHARED LINKAGE TO PHYREV V135 P0316
J001 V070 P0460

Example 9c. Use of share b.

s phyrev 135
f share b phyrev 135 247
f share b phyrev 135 582
o p i t a linkage
g

PHYSICAL REVIEW
VOLUME 135

J001 V135 P0247

ELECTRON SPIN DOUBLE RESONANCE STUDIES OF F CENTERS IN KCL. I

MORAN P. R.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460	J001 V091 P1071	J001 V098 P1787
J001 V102 P0151	J001 V110 P0630	J001 V114 P1245
J001 V115 P1506	J001 V118 P1024	J001 V124 P0442
J011 V022 P0989	J011 V026 P1124	J011 V029 P1692
J030 V026 P0167	J031 V032 P0775	J046 V005 P0183
J052 V008 P0299		

J001 V135 P0316

DOUBLE-RESONANCE PHENOMENA IN THE GASEOUS LASER

CULSHAW W.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460 J030 V026 P0167

SHARED LINKAGE TO PHYREV V135 P0582

J001 V076 P0833 J001 V107 P1559 J096 V229 P1213

J001 V135 P0470

SPIN-LATTICE RELAXATION OF F CENTERS IN KCL. ISOLATED F CENTERS

FELDMAN D. W.

WARREN R. W.

CASTLE J. G., JR.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V091 P1071

J001 V135 P0582

SPIN RELAXATION OF OPTICALLY PUMPED CESIUM

FRANZ F. A.

LUSCHER E.

SHARED LINKAGE TO PHYREV V135 P0582

J001 V076 P0833	J001 V098 P0478	J001 V105 P1487
J001 V107 P1559	J001 V108 P1453	J001 V115 P0850
J001 V123 P0544	J001 V132 P0712	J003 V067 P0853
J012 V036 P0135	J012 V037 P2504	J017 V031 P0986
J031 V034 P0589	J034 V176 P0045	J038 V011 P0255
J041 V001 P0052	J041 V001 P0054	J041 V005 P0373
J045 V047 P0460	J046 V003 P0009	J046 V003 P0372
J046 V008 P0009	J046 V008 P0529	J046 V009 P0011
J049 V007 P0277	J052 V049 P0127	J074 V028 P0646
J096 V229 P1213	J096 V241 P0865	J096 V246 P1522
J096 V254 P3829	J256 V004 P0177	J273 V006 P1148

Example 9c continued on next page

Example 9c continued from page 17

J001 V135 P0591
STUDY OF SPIN-EXCHANGE COLLISIONS IN VAPORS OF RB85, RB87, AND
CS133 BY PARAMAGNETIC RESONANCE

MOOS H. WARREN

SANDS RICHARD H.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

SHARED LINKAGE TO PHYREV V135 P0582

J041 V001 P0054

J001 V135 P0727

LINE SHAPES OF PARAMAGNETIC RESONANCES OF CHROMIUM IN RUBY

GRANT W. J. C.

STRANDBERG M. W. P.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V091 P1071

J001 V135 P1046

FAST-PASSAGE EFFECTS IN THE NUCLEAR MAGNETIC RESONANCE OF FE57
IN PURE IRON METAL

COWAN DAVID L.

ANDERSON L. WILMER

SHARED LINKAGE TO PHYREV V135 P0247

J001 V091 P1071

J001 V135 P1068

SPIN-LATTICE RELAXATION IN FREE-RADICAL COMPLEXES

KRISHNAJI

MISRA B. N.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

J001 V135 P1099

LOW-FIELD RELAXATION AND THE STUDY OF ULTRASLOW ATOMIC MOTIONS
BY MAGNETIC RESONANCE

SLICHTER CHARLES P.

AILLON DAVID

SHARED LINKAGE TO PHYREV V135 P0247

J001 V098 P1787

J001 V135 P1498

FORCED TWO-LEVEL OSCILLATOR

SENITZKY I. R.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

J001 V135 P1622

LATTICE SUM EVALUATIONS OF RUBY SPECTRAL PARAMETERS

ARTMAN J. O.

MURPHY JOHN C.

SHARED LINKAGE TO PHYREV V135 P0582

J046 V008 P0529

end of Example 9c.

Example 9d. Two share b's "and"ed. (Compare with 9c)

s phyrev 135

f share b j1 135 247 and share b j1 135 582

o p i t a linkage

g

PHYSICAL REVIEW

VOLUME 135

J001 V135 P0316

DOUBLE-RESONANCE PHENOMENA IN THE GASEOUS LASER

CULSHAW W.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460 J030 V026 P0167

SHARED LINKAGE TO PHYREV V135 P0582

J001 V076 P0833 J001 V107 P1559 J096 V229 P1213

J001 V135 P0591

STUDY OF SPIN-EXCHANGE COLLISIONS IN VAPORS OF RB85, RB87, AND
CS133 BY PARAMAGNETIC RESONANCE

MOOS H. WARREN

SANDS RICHARD H.

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

SHARED LINKAGE TO PHYREV V135 P0582

J041 V001 P0054

Example 9e. Use of "but not" with share b. (Compare with 9c)

f share b j1 135 247 but not title +resonance

o p i t linkage

g

PHYSICAL REVIEW

VOLUME 135

J001 V135 P0470

SPIN-LATTICE RELAXATION OF F CENTERS IN KCL. ISOLATED F CENTERS

SHARED LINKAGE TO PHYREV V135 P0247

J001 V091 P1071

J001 V135 P1068

SPIN-LATTICE RELAXATION IN FREE-RADICAL COMPLEXES

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

J001 V135 P1498

FORCED TWO-LEVEL OSCILLATOR

SHARED LINKAGE TO PHYREV V135 P0247

J001 V070 P0460

OUTPUT PRINT

This command selects the information to be associated with the found item and displays it to the user. The information to be displayed may be any combination of identification (i), title (t), authors (a), location (l), citations (c) and when applicable, linkage.

The word "all" is an acceptable descriptor for search find and output print.

```
s all
find all
o print all
```

will print the entire TIP library. This exercise is not recommended to the casual user. However

```
s phyrev 135
find all
o print all
```

will print all the information in v 135 of Physical Review. See page 10, example 4 for an illustration of "output print all".

OUTPUT SAVE

So far the only output used has been "output print". With this command the information found and printed is erased immediately after print and is not available for further use. It is frequently desirable to save a file or parts of it for subsequent search, print or retrieval. This may be accomplished by the "output save" command as illustrated in example 10, where we search the library for titles that contain the word maser. The entire data on each article is saved and filed under the heading "maser titles".

Example 10. Use of "output save".

```
s all
f t +maser
o save all
name save file maser titles
go
```

ANNALS OF PHYSICS

```
VOLUME 26
VOLUME 27
VOLUME 28
VOLUME 29
VOLUME 30
VOLUME 31
VOLUME 32
VOLUME 33
VOLUME 34
```

APPLIED PHYSICS LETTERS

```
VOLUME 4
VOLUME 5
VOLUME 6
VOLUME 7
```

Notes on Example 10:

1. The illustration in example 10 was cut off before completion. Ordinarily the computer will print each journal name and volume as it is being searched and end with the standard statement of search completed.
2. In the present configuration of TIP, with 20,000 articles in store, "search all" takes of the order of three minutes. This is a sizeable work load for MAC and will therefore be given a low priority. Users are advised not to attempt "search all" in the daytime when the call on MAC is very heavy. Weekends and nights are more suitable. The user must also weigh the time cost of this procedure against his total time allotment.
3. The file is saved in the users track allotment. Users who create sizeable files should make sure that enough tracks are available to them. As a rule of thumb, one may consider that "save all" will require about 50 tracks for 200 articles.
4. Saved files are not part of the TIP library and are not available outside the program number in which they were created. They may be linked to other users through the standard link procedure. Users who create special files that may be of general use to others are requested to communicate with TIP personnel for possible inclusion in the common TIP library.

Example 11. Using the TIP program on a saved file.

search maser titles
find title hydrogen
o p i t
g

MASERTITLES
J646 V004 P0122
CAVITY TUNING AND PRESSURE DEPENDENCE OF FREQUENCY IN THE
HYDROGEN MASER

J646 V006 P0034
HYDROGEN MASER AND CESIUM BEAM TUBE FREQUENCY STANDARDS
COMPARISON

J646 V006 P0035
HYDROGEN MASER FREQUENCY COMPARISON WITH SWISS CESIUM ATOMIC
BEAM STANDARD

J001 V137 P1621
SPIN EXCHANGE AND SURFACE RELAXATION IN THE ATOMIC HYDROGEN
MASER

J001 V138 P0972
HYDROGEN-MASER PRINCIPLES AND TECHNIQUES

Example 12. A list of all Physical Review articles in the "maser titles" file.

s maser titles
f i phyrev
o p t i
g

MASERTITLES

J001 V130 P0675
THEORY OF CAVITY MASERS

J001 V130 P0806
QUANTUM STATISTICS OF MASERS AND ATTENUATORS

J001 V131 P0095
MECHANISM OF SECOND HARMONIC GENERATION OF OPTICAL MASER BEAMS
IN QUARTZ

J001 V131 P2038
MASER OSCILLATIONS AT 0.9 AND 1.35 MICRONS IN $\text{CaWO}_4:\text{Nd}^{3+}$

J001 V133 P1221
TEST OF SPECIAL RELATIVITY OR OF THE ISOTROPY OF SPACE BY USE
OF INFRARED MASERS

J001 V133 P1244
OPTICAL MASER ACTION IN C, N, O, S, AND BR ON DISSOCIATION OF
DIATOMIC AND POLYATOMIC MOLECULES

J001 V133 P1476
NOBLE GAS OPTICAL MASER LINES AT WAVELENGTHS BETWEEN 2 AND 35
M

J001 V134 P0299
THEORY OF PHONON-TERMINATED OPTICAL MASERS

J001 V134 P1429
THEORY OF AN OPTICAL MASER

J001 V135 P0543
FREQUENCY PUSHING AND FREQUENCY PULLING IN A HE-NE GAS OPTICAL
MASER

J001 V137 P1621
SPIN EXCHANGE AND SURFACE RELAXATION IN THE ATOMIC HYDROGEN
MASER

J001 V138 P0972
HYDROGEN-MASER PRINCIPLES AND TECHNIQUES

J001 V139 P0617
MODE COMPETITION AND FREQUENCY SPLITTING IN
MAGNETIC-FIELD-TUNED OPTICAL MASERS

J001 V139 P0635
THEORY OF A TRAVELING-WAVE OPTICAL MASER

SEARCH COMPLETED, 60 ARTICLES.
1.58 SECONDS, 37.9 ARTICLES/SEC.
14 ARTICLES FOUND.

OUTPUT STORE

The library information on the discs is compressed into a compact form in order to conserve tracks. Before this information can be displayed to the user, it must be unpacked to its original format. The "output print" command unpacks the data before display. The command "output save" creates a file in the packed mode which is subsequently searched and unpacked by the print command. One cannot print any file be it TIP or saved and display it in intellegible form without requesting "output print" through the console. This introduces an inconvenience when a massive print out is wanted that would take too long to accomplish on the console. When large amounts of print out are expected, use the command "output store". The command will unpack this information and store it in the same format as "output print". The user may then request off-line printing of the file on the fast printers at MAC.

Example 13. How to use "store file" and "request print".

```
search maser titles
find all
output store all
name store file maser store
*o
```

MASERTITLES

```
SEARCH COMPLETED,      60 ARTICLES.
 13.54 SECONDS,        4.4 ARTICLES/SEC.
 60 ARTICLES FOUND.
```

```
QUIT,
R 20.400+2.533
```

```
rquest print maser store
W 1129.3
R .750+.383
```

Notes on Example 13:

1. The stored files, like the saved files, are on the user's tracks and are available only through the program number in which they were created.
2. "Request print" is not part of the TIP program. The user must, therefore, QUIT out of TIP after "store file" is completed and then execute the off-line print request.
3. For full instructions on "request print" see TSS manual section AH.6.06.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL INFORMATION PROGRAM*

TIP LIBRARY MAINTENANCE

A description of the methods and operating procedures
used by TIP personnel in the production of a machine-usable library
of journal articles, its maintenance and up-dating.

M. M. Kessler
October, 1967

*Supported by a grant from the National Science Foundation

TIP LIBRARY MAINTENANCE

The creation of an appreciable body of literature in machine-usable form, its organization into memory files and a monitor system that will detect flaws and check deterioration is itself a problem for research and development. We presently maintain a library of roughly 100,000 articles from 38 physics journals. No matter what happens to the TIP system as a whole, it is clear that the TIP library is an important asset that must be maintained and given continuity.

We presently process about 1500 articles each month. This is accomplished with a staff consisting of four Flexowriter operators, two proof-readers, a data processing assistant, and a professional programmer who acts as supervisor. The operation also involves considerable use of the computer for formatting and editing. We regard our present input operation as an adequate beginning in need of a serious engineering effort that will make the flow of data smoother and more economical.

The TIP package, consisting of the machine-usable library and program files, represents a commitment to the public that certain minimal facilities will be available and functioning at all times and for every member of the MAC community of some 200 consoles and 500 subscribers. It is in the nature of a public utility that guarantees an average level of service. In order to insure this level of service, a system of up-dating and maintenance had to be developed that must be in constant operation, largely unknown to the user, but always on the alert to catch and identify errors and break-downs. Indeed, problems concerning efficient work flow and maintenance of system integrity are themselves important areas of development in a prototype system.

Like the iceberg, nine-tenths of which is submerged and invisible, a large and significant part of the TIP effort has been devoted to the maintenance of the system, an effort whose chief virtue is that it be unknown to the user of the

system but always present in the background to insure its proper functioning.

This aspect of TIP involves three sections:

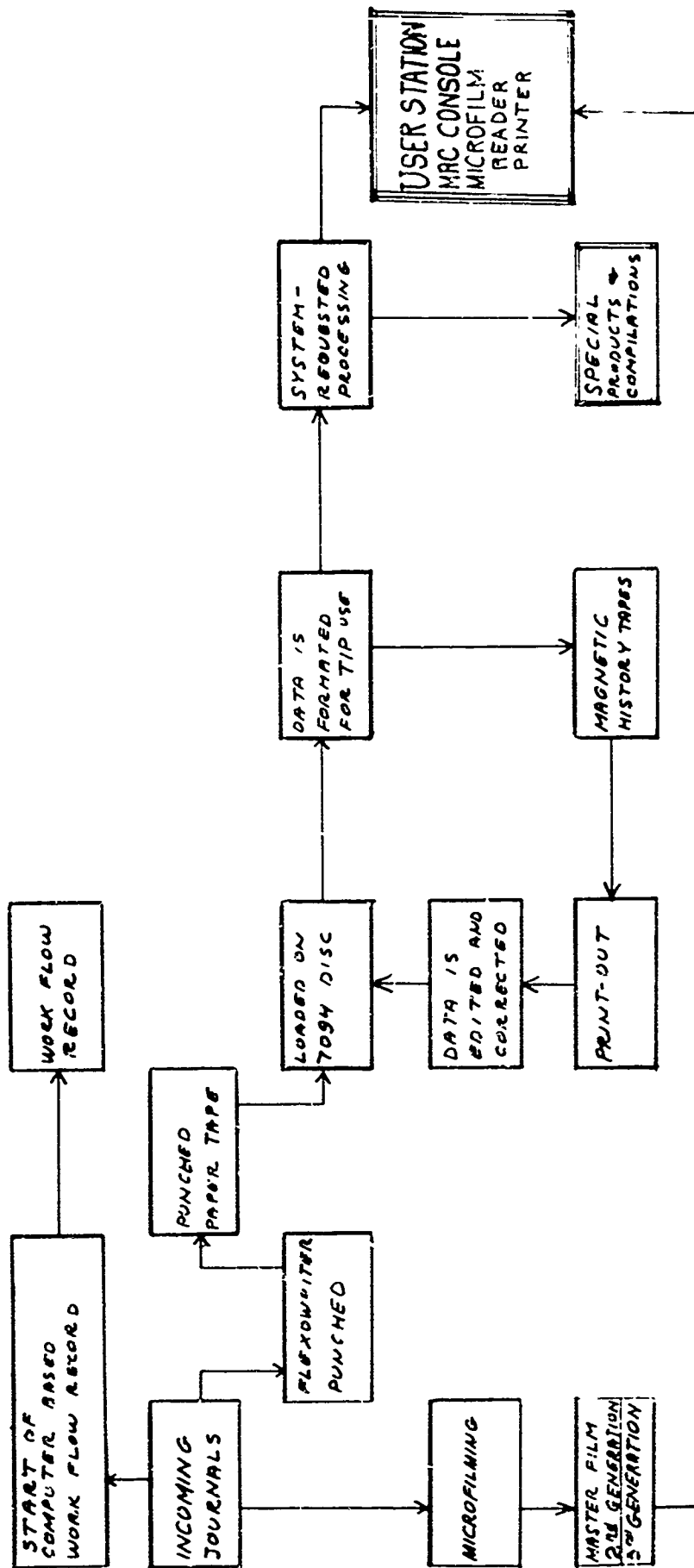
- a. The flow of information from journal to TIP public library.
- b. Check points and monitors for disc inventory.
- c. Internal communications and file management.

Fig. 1 is a simplified chart showing the flow of information from the time the journal is received to its final state of availability on disc for TIP search. TIP subscribes to 38 journals (see List A) which have been chosen by the criterion of the bibliographic matrix previously described.¹ These journals make up a monthly contribution of some 1500 articles. Depending on the criteria used, we estimate that our library represents some 50 to 75% of the working physicists' current journal literature.

As each issue arrives, the proper inventory procedure is initiated that follows the issue step by step through its processing cycle. This constitutes the work flow record that is cumulated and examined periodically for bottle necks, operator efficiency, and status reports.

The contents page is reproduced and sent to the keypunch room where a record is generated of the journal, volume, issue number, date, and the beginning page number of each article in the issue. This information will be kept as a separate file and used periodically as a check to certify the integrity of the disc data. At the same time, the entire issue is sent to be keypunched. Until the end of 1966, the data input was from IBM cards. Fig. 2 shows a sample of the pre-1967 data input content and format. The article identification was in terms of a number code for the journal name, volume, and page. The title, author, and institutional location were punched uncoded. The bibliography of each paper was again coded by journal, volume, and page. Note that in the pre-1967 data, only journal articles were fully identified in the bibliography. References to books, reports, theses, etc. were mentioned by type only, to give statistical information, but were not fully specified.

¹ M. M. Kessler, "The MIT Technical Information Project". Physics Today, March 1965, pp. 28-36.



TIP INPUT PROCEDURE

FIG. 1

LIST A

LIST OF JOURNALS PROCESSED FOR TIP

Annals of Physics (New York)
Applied Physics Letters
Canadian Journal of Physics
* Chinese Journal of Physics
* Discussions of the Faraday Society
Helvetica Physica Acta
* Indian Journal of Physics
Japanese Journal of Applied Physics
* JETP Letters
Journal of Applied Physics
Journal of Chemical Physics
* Journal of Chemical Physics (Supplement)
* Journal of Mathematical Physics
* Journal of the Optical Society of America
Journal of the Physical Society of Japan
* Molecular Physics
Nuclear Physics (Series A)
* Nuclear Physics (Series B)
Nuovo Cimento (Series A)
* Nuovo Cimento (Series B)
* Nuovo Cimento (Supplemento)
* Philosophical Magazine
Physica
Physical Review
Physical Review Letters
Physics Letters (Series A)
* Physics Letters (Series B)
Physics of Fluids
Proceedings of the Physical Society
* Proceedings of the Royal Society (London)
Progress of Theoretical Physics
* Progress of Theoretical Physics (Supplement)
* Soviet Journal of Nuclear Physics
Soviet Physics - JETP
* Soviet Physics - Semiconductors
Soviet Physics - Solid State
Soviet Physics - Technical Physics
* Transactions of the Faraday Society

* Added to TIP library in 1966

TIP DATA FORMAT
PRIOR TO 1967

search phyrev 135
find identification j1 135 960
output print all
go

PHYSICAL REVIEW
VOLUME 135
J001 V135 P0960
VIBRATIONAL AND CENTRIFUGAL EFFECTS ON THE MAGNETIC
SUSCEPTIBILITY AND ROTATIONAL MAGNETIC MOMENT OF THE HYDROGEN
MOLECULE

CHAN SUNNEY I.
PASADENA, CALIFORNIA
CALIFORNIA INSTITUTE OF TECHNOLOGY
GATES AND CRELLIN LABORATORIES OF CHEMISTRY
KENBERRY DENNIS
DAS T. P.

RIVERSIDE, CALIFORNIA
UNIVERSITY OF CALIFORNIA
DEPARTMENT OF PHYSICS

J001 V034 P0057	J001 V041 P0713	J001 V041 P0721
J001 V058 P0310	J001 V078 P0711	J001 V080 P0476
J001 V085 P0937	J001 V087 P1075	J001 V094 P0350
J001 V094 P0893	J001 V103 P1254	J001 V111 P0203
J001 V112 P1929	J001 V115 P0897	J001 V126 P0146
J002 V000 P0000	J002 V000 P0000	J003 V065 P0178
J012 V009 P0061	J012 V019 P1030	J012 V020 P0527
J012 V021 P2070	J012 V023 P1131	J012 V032 P0105
J012 V035 P1065	J012 V035 P1967	J012 V037 P0214
J012 V037 P1527	J012 V038 P1263	J027 V000 P0000
J027 V000 P0000	J030 V032 P0231	J030 V035 P0130
J046 V006 P0019	J055 V035 P0730	J160 V004 P0061
J311 V003 P0017	J311 V010 P0278	

Fig. 2

Thus in the example shown in Fig. 2, the sixth line in the bibliography, "J002 V000 P000", means that a book was cited in this article, but no further identification of the book is given. "J032 V000 P000" refers to the citation of a report, while "J033" refers to a thesis. In the above cases, the citations are merely identified for statistical purposes but are not given in full.

Each keypunch operator was provided with a table of journal titles and the code number for each title. If a new journal title was encountered, the operators were instructed to make a special card with the new title. These were examined periodically and each time a new title appeared in appreciable numbers, it was assigned a number to be used henceforth.

This format was acceptable as long as TIP was a purely experimental system intended to be used for test and trial purposes only. In the period of 1965-1966, several conversations with members of the AIP and NSF were held concerning the possible expansion of TIP into a prototype system. It was agreed that the AIP could make better use of our data if considerable expansion were undertaken, both in the number of journals covered and in the nature of the data recorded from each article.

In response to these verbal understandings, TIP increased its journal coverage from the original 20 journals to its present list of 38 journals. We also incorporated many changes in the input format and content.

Fig. 3 shows the new data format. Fig. 4 is the same data expanded into a more readable format for editing and corrections. The main operational difference introduced in January, 1967 is the use of a paper tape for input instead of punched cards. The most significant contribution in this shift is a much expanded keyboard, particularly the availability of upper and lower case letters, as well as certain symbols and special codes. We have also made changes and expanded the record, particularly with regard to the references. Instead of using our private list of code numbers for journal names, we have adopted the CODEN nomenclature. A sample list of journal names, their CODEN abbreviations, and the equivalent TIP code number is shown in Fig. 5. Each journal is listed not only by its correct full title but also by the numerous abbreviations that are frequently encountered in the literature. The references are identified by author

TIP DATA SAMPLE: POST-1967

- *J APNY
- *V 41
- *P 1
- *T General Nondynamical Formalism for Reactions with Particles of Arbitrary Spin. Number of Form Factors. Parity Conservation.
- *A Paul L. Csonka, Michael J. Moravcsik, and Michael D. Scadron
- *L Lawrence Radiation Laboratory, University of California, Livermore, California
- *B #A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, #J PHLT #V 15, #P 353, #D (1965)
- *B #A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, #J PHLT #V 14, #P 861, #D (1956)
- *B #A P. L. Csonka, J. J. Moravcsik, and M. D. Scadron, #J Ann. Phys., in press, and UCRL-14222
- *B #A M. J. Moravcsik, #T Proc. Intern. Conf. Polarization Phenomena Nucleons, 1965, Karlsruhe, in press
- *B #A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, #J NUCI #V 42, #P 743, #D (1966)
- *B #A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, #J PHRV #V 143, #P 775, #D (1966)
- *B #A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron, #J PHLT #V 20, #P 67, #D (1966)
- *B #A M. J. Moravcsik, "The Non-Dynamical Structure of Particle Reactions," Lectures delivered at the 1st Pacific International Summer School in Physics, Honolulu, August 1965. In "Recent Developments in Particle Physics" Gordon and Beach, New York, in press.
- *B #A P. L. Csonka, #J RMPH #V 37, #P 177, #D (1965)

- *P 52
- *T The Normal Modes of Oscillation of a Uniformly Charged Drop about Its Saddle-Point Shape
- *A James Rayford Nix
- *L Lawrence Radiation Laboratory, University of California, Berkeley, California
- *B #A D. L. Hill and J. A. Wheeler, #J PHRV #V 89, #P 1102, #D (1953)
- *B #A U. L. Businaro and S. Gallone, #J NUCI #V 1, #P 629, #D (1955)
- *B #A J. R. Nix and W. J. Swiatecki, #J NUPY #V 71, #P 1, #D (1965)
- *B #A J. R. Nix, Ph.D. Thesis, University of California Lawrence Radiation Laboratory Report UCRL-11338, April 1, 1964 (unpublished)
- *B #A Lawrence Willets, "Theories of Nuclear Fission," Chap. 3, Subsec. 4.2, pp. 50-53. #PN Clarendon Press, #PL Oxford, #PD 1964)
- *B #A J. J. Griffin, Paper No. SM-60/4, in "Physics and Chemistry of Fission," Proceedings of a Symposium held in Salzburg, 1965, Vol. I, p. 23 (International Atomic Energy Agency, Vienna, 1965)
- *B #A W. D. Myers and W. J. Swiatecki, #J NUPH #V 81, #P 1, #D (1966)
- *B #A W. J. Swiatecki, #J PHRV #V 104, #P 993, #D (1956)
- *B #A S. Cohen and W. J. Swiatecki, #J APNY #V 22, #P 406, #D (1963)
- *B #A Horace Lamb, "Hydrodynamics," 6th ed., Sec. 9, pp. 6-8; Sec. 45, pp. 47-48. Dover, New York, 1945.
- *B #A Herbert Goldstein, "Classical Mechanics," Chap. 1, Sec. 6, pp. 22, 23. Addison-Wesley, Reading, Mass, 1959.
- *B #A N. Bohr and J. A. Wheeler, #J PHRV #V 56, #P 426, #D (1939)
- *B #A W. J. Swiatecki, Paper No. P/651, in Proc. Second U. N. Intern. Conf. Peaceful Uses Atomic Energy, Geneva, 1958, Vol. 15, "Physics in Nuclear Energy," p. 248 (United Nations, Geneva, 1958)
- *B #A S. Cohen and W. J. Swiatecki, #J SPNY #V 19, #P 67, #D (1962)
- *B #A A. R. Edmonds, "Angular Momentum in Quantum Mechanics," Chap. 3, Sec. 7, pp. 45-50; Chap. 4, Sec. 6, pp. 62, 63. Princeton Univ. Press, Princeton, New Jersey, 1957.

Fig. 3

TIP DATA SAMPLE - EXPANDED: POST-1967

*P 1

*T General Nondynamical Formalism for Reactions with Particles of Arbitrary Spin. Number of Form Factors. Parity Conservation.

*A Paul L. Csonka, Michael J. Moravcsik, and Michael D. Scadron

*L Lawrence Radiation Laboratory, University of California, Livermore, California

*B

#A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron,

#J PHLT

#V 15,

#P 353,

#D (1965)

*B

#A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron,

#J PHLT

#V 14,

#P 861,

#D (1956)

*B

#A P. L. Csonka, J. J. Moravcsik, and M. D. Scadron,

#J Ann. Phys., in press, and UCRL-14222

*B

#A M. J. Moravcsik,

*T Proc. Intern. Conf. Polarization Phenomena Nucleons, 1965, Karlsruhe, in press

*B

#A P. L. Csonka, M. J. Moravcsik, and M. D. Scadron,

#J NUCI

#V 42,

#P 743,

#D (1966).

*B

#A M. J. Moravcsik, "The Non-Dynamical Structure of Particle Reactions," Lectures delivered at the 1st Pacific International Summer School in Physics, Honolulu, August 1965. In "Recent Developments in Particle Physics" Gordon and Beach, New York, in press.

*B

#A P. L. Csonka,

#J RMPH

#V 37,

#P 177,

#D (1965)

Fig. 4

CODEN LIST USED BY TIP

BJCA 891	BRIT. J. CANCER.
BCEL 885	BRITISH COMMUNICATIONS AND ELECTRONICS
BJAP 324	BRITISH JOURNAL OF APPLIED PHYSICS
BUPS 901	BULL. ACAD. OF SCIENCE PHYS. SER.
BARM 882	BULL. ACAD. POL. SCI. SER. SCI. MATH. ASTRON. PHYS.
BARC 881	BULL. ACAD. POLON. SCI. SER. SCI. CHIM.
BAPM 880	BULL. ACAD. POLON. SCI., SER. SCI. MATH., ASTR. ET PHYS.
BAMI 222	BULL. AM. METEOROL. SOC.
BAPS 046	BULL. AM. PHYS. SOC.
BAPS 046	BULL. AMER. PHYS. SOC.
BKNS 892	BULL. BORIS KIDRICH INST. NUCL. SCI.
BCSJ 539	BULL. CHEM. SOC. JAPAN
BCSA 775	BULL. CLASSE SCI. ACAD. ROY. BELG.
BUGU 899	BULL. GEOL. INST. UNIV. UPPSALA
BUGH 483	BULL. GEOL. SOC. AM.
BICR 886	BULL. INST. CHEM. RES. KYOTO UNIV.
BULM 900	BULL. INST. MET.
BKIP 744	BULL. KOBAYASHI INST. PHYS. RES.
BUDY 898	BULL. NANIWA UNIV.
BRCI 894	BULL. RES. COUNCIL ISRAEL
BRCI 894	BULL. RESEARCH COUNCIL ISRAEL
BUUN 567	BULL. SCRIPPS INST. OCEANOGR.
BSSA 897	BULL. SEISMOL. SOC. AMER.
BSCF 896	BULL. SOC. CHIM. (FRANCE)
BSBQ 795	BULL. SOC. CHIM. BELG.
BSBQ 795	BULL. SOC. CHIM. BELGE
BSCF 896	BULL. SOC. CHIM. FR.
BUFC 135	BULL. SOC. FR. MIN. CRIST.
BUFC 135	BULL. SOC. FRANC. MINERAL. CRIST.
BSRS 354	BULL. SOC. ROY. SCI. LIEGE
BUTY 902	BULL. TOKYO INST. TECH., SERIES B
BSBQ 795	BULLETIN DE LA SOCIETE CHIMIQUE DE BELGIQUE
BSRS 354	BULLETIN DE LA SOCIETE ROYALE DES SCIENCES DE LIEGE
BUPS 901	BULLETIN OF THE ACADEMY OF SCIENCE OF THE U.S.S.R. PHYS. SER.
BAMI 222	BULLETIN OF THE AMERICAN METEOROLOGICAL SOCIETY
BAPS 046	BULLETIN OF THE AMERICAN PHYSICAL SOCIETY
BCSJ 539	BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN
BUGH 483	BULLETIN OF THE GEOLOGICAL SOCIETY OF AMERICA
BKIP 744	BULLETIN OF THE KOBAYASHI INSTITUTE OF PHYSICAL RESEARCH
CAPH 174	CAHIERS DE PHYS.
CAPH 174	CAHIERS DE PHYSIQUE
CAPH 174	CAHIERS PHYS.
CJCH 361	CAN. J. CHEM.
CJCE 912	CAN. J. CHEM. ENG.
CJPH 055	CAN. J. PHYS.
CJRE 009	CAN. J. RES.
CJTE 913	CAN. J. TECH.
CMBU 914	CAN. MATH. BULL.
CJPH 055	CANADIAN J. OF PHYSICS
CJCH 361	CANADIAN JOURNAL OF CHEMISTRY
CJPH 055	CANADIAN JOURNAL OF PHYSICS
CRBN 919	CARBON
CEAG 280	CER. AGE
CEAG 280	CERAMIC AGE
CHAB 909	CHEM. ABS.
CHAB 909	CHEM. ABSTR.
CHBE 301	CHEM. BER.
CENE 395	CHEM. ENG. NEWS
CEPR 495	CHEM. ENG. PROGR.
GESC 908	CHEM. ENG. SCI.
GESC 908	CHEM. ENNG SCI.

Fig. 5

and year as well as by their journal, volume, and page. This is a redundancy that was added to the post-'967 data in response to numerous requests from both users and librarians. A sizeable expansion of the keypunching load was introduced by our decision to record the full identification of all references, including books, reports, etc. (see Fig. 4), whereas in the old format such material was merely noted without being fully recorded for bibliographic retrieval. These changes required a considerable modification in the TIP program to make it function on both sets of data without any interference to the user.

The paper tape is loaded onto the disc and presumably need not ever be used again because all subsequent manipulations, corrections, etc. are done on the computer. We, nevertheless, keep the punched paper tape as an ultimate insurance against disaster or loss. A formatting program is applied to the disc-recorded paper tape image that puts the information into TIP-usable format. This is largely a matter of internal coding and compression designed to save memory space on the disc and processing time in subsequent handling. At this point, a magnetic tape copy is made of the compressed and coded disc record. This tape, a primary history record of the TIP public library, is kept as further insurance against accidental loss or misplacement of disc files. A printout from the history tape provides the text for editing. All corrections are noted on the printed paper and executed from the console on the disc record. The cycle may then be repeated until the printed record is deemed correct. In actual practice, we edit and correct just once. This input process is done weekly on the keypunched record as it accumulates. The disc record of the weekly input is then processed for various current outputs and merged with the previous data. Thus, a new issue will join the previous issues of the given journal until the volume is completed. At the end of the processing and editing cycle, the material is available to the users at the MAC consoles and the various TIP search and retrieval programs will be applicable. It is also available in other formats on tape cards, paper tape, or printout.

At this time, we should mention an auxiliary process functionally interlocking with TIP but not an integral part of TIP. It has long been recognized by us that with the final output of TIP, being limited to title, author, and

bibliographic information, remains open-ended until the user has consulted the actual text of the paper that has been suggested by a TIP search procedure. Certain experiments and search strategies depend heavily on user feed-back on the value and relevance of the search. This feed-back must be based on the user's judgment of the contents of papers retrieved by TIP and not on title and author only. An examination of the technical problems involved in text retrieval convinced us that at this time, it would be quite impractical to store any appreciable amount of text on the disc in digital form. Even limited text, such as abstracts or enriched titles, would unduly load our memory capacity. It is not the intent of TIP to develop and pioneer in the area of text storage and retrieval. We have, therefore, taken advantage of commercially available microfilm equipment as an instrument for text availability and have adapted it for use at several TIP stations.

When the keypunching operation is finished, the journal-issue is sent to our microfilm laboratory where it is photographed on 16 mm. film and placed in 100-foot reels. One hundred feet of film is a convenient unit that will accommodate one volume of most journals. The resulting unit of film is a cartridge for each journal-volume which can be searched, read, and copied on one of several commercially available reader-printer machines. Two such reader-printer units have been installed at two MAC stations, one at the central TIP installation in the MIT Library, the other in a room available to graduate students. A user at one of these stations may consult TIP in one of the various modes available for search and retrieval. The output of the system will consist of a set of papers identified by journal, volume, page, title, and author. Within reach of the user is the full text of the paper which may be immediately consulted. In this way, the user can judge the relevance of TIP results and on this basis modify his search strategy at the console.

The previous paragraphs described the progress of each issue of a journal from arrival to the time when the pertinent information is located on the disc and ready for use by TIP subscribers. This is by no means our last preoccupation with the integrity of the data. A program of checks and monitors exists that alerts TIP operators to faults in the data bank. These faults may

be massive or trivial; they may be due to errors in the initial punching and input or to computer failures and accidental erasures. At any rate, inventory control must be an on-going process, and facilities must be provided to correct the faulty data. Several levels of certification exist. Each of these is monitored and reassessed from time to time against a carefully constructed inventory file.

a. Journal certification.

A file exists of all the volumes of each journal in the TIP library.

There is a weekly check of the library against this file. If volumes are found to be missing, they are replaced. This is a rare occurrence.

b. Issue certification.

The date, beginning and last page of each constituent issue of a volume are checked against a file.

c. Article certification.

A check is made periodically of the entire library against the beginning page numbers of each article as they were previously and independently recorded.

d. Field certification.

Each article is checked to make sure that each of the five fields, i.e. identification, title, author, location, and bibliography, exists.

e. Text certification.

No automatic program has been established for the periodic check of the accuracy of the text. Prior to 1967, each item was keypunched and verified. After 1967, we have established a procedure of proofreading and correction from the console. Once the text is proofread and corrected, we assume that the errors are minimal and acceptable. Our experience with daily use of the system confirms the decision that one proofreading is acceptable.

A massive damage to the data of type "a", where entire volumes of journals disappear, comes about only as a result of a major break-down in the computer such as a power failure or mechanical damage to the discs. To insure against such total failures, the entire record of the MAC disc is copied on tape

on tape each day and kept for a period of six months. With this arrangement, one is insured against the loss of more than a day's accumulation of data.

A more local or partial damage to the TIP library is corrected from the TIP history tapes that we accumulate as a result of our input procedure. As a final insurance, we keep the original cards and paper tape that contain the raw data.

In addition to the TIP files of journal articles, there also exists on the disc a large number of programs, memos, partial results, and other files of data and programs in various states of completion and belonging to the various members of the TIP staff. These must be under constant monitor and surveillance. A monitor program exists that gives a daily listing of all the files in existence, their names and the name of the person who created them. This allows the designated file supervisor to make sure that the system's memory is not cluttered with old and useless material. It also keeps the staff informed of the programs that are currently available or that are being worked on at a given time. Programs and other files that are not in current operation are removed from the disc and placed on reels of magnetic tape. These tapes together with the history tapes previously mentioned make up a reserve tape library of data and programs that can be called on by the staff when needed. The naming, indexing, and retrieval of tapes from this library is an important problem in large system management and, indeed, is not unlike the more general retrieval problem that TIP handles. Monitoring of the data and program files contributes to the management and usability of the system. In order to have an overall picture of how and by whom the system is being used, we also monitor and record all users by name, console location, the type of inquiry or engagement called for, and the total time spent within the TIP system. This information is listed weekly and provides a rough measure of system utilization by TIP staff as well as by outside on-line users.

One more monitoring effort must be mentioned which will become increasingly important when costs of service, maintenance, and development will have to be considered. Project TIP now has seven professional staff members, six part-time students and five non-professional employees. Each has a certain amount of time and memory space assigned to him. In addition to personal assign-

ments, there is also some time and memory assigned to certain functions or projects within TIP independent of the operators. It is most significant to gather information on the distribution of the total TIP facilities among the various people and projects that constitute the entire effort. Some break-down is needed on the cost of data input, program writing, training of new staff, the production of certain services, etc., etc. In order to gather this information, a program exists that gives us a weekly printout of the distribution of TIP facilities among the various categories of users, staff, and functions.

The management and maintenance of the internal housekeeping of the system, even at our prototype size, is a significant overhead item. We have, nevertheless, developed this aspect of the system because we feel that a realistic prototype must include within it the elements of management, control, cost accounting, inventory checks, etc.

A Bibliographic Search by Computer

Updating plasma-physics data was a chance to experiment with information and programs of the Technical Information Project at MIT. The computer searched for indicative words in titles, papers that shared bibliographical references and those that referred to papers that have become classics in plasma physics.

by Sanborn C. Brown

ANY COLLECTION OF DATA one can make these days is out of date before it is published. This is true, for example, of my *Basic Data of Plasma Physics*.¹ In the book I tried to bring together in useful form the data of gaseous electronics and plasma physics upon which scientists base calculations and further work with basic parameters.

Updating this book seemed an ideal computer experiment. I used the Information Retrieval Service of the TIP (Technical Information Project) program not only to find material that has come out since publication of the book but also to ar-

range the program so that the computer could continue in the future to retrieve relevant information.

Myron Kessler described TIP last year in *PHYSICS TODAY*.² My experiment illustrates a particular use of the physics literature programed into the project and shows some things a computer can do as a reference and bibliographical tool.

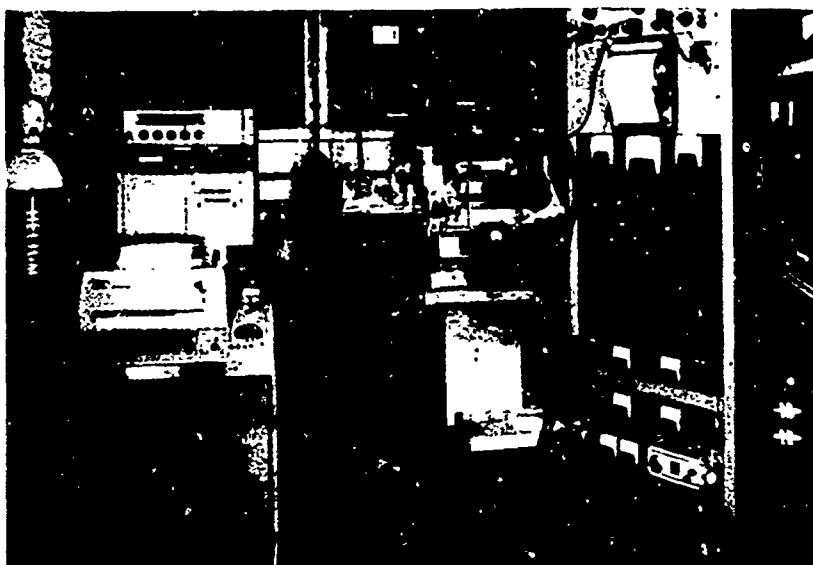
There are, of course, many possible and useful permutations and combinations of both the library program and the methods of searching. I will not try to present any kind of a definitive treatment but merely to call attention to this most useful and tremendously time-saving method of trying to keep up with the modern flood of physics literature.

Material available from TIP

The Technical Information Project, upon which this experiment is based, has programed 25 physics journals from the past few years and *The*

The author, who was born in Beirut, Lebanon, has spent most of his adult life studying and teaching physics at Dartmouth College and MIT. The war years of 1941-44 he spent at the US Office of Scientific Research and Development. His interests, in addition to plasma physics, include atoms, molecules, and history and philosophy of the physical sciences.





TELETYPE CONSOLE used for bibliographic search connects with Project MAC time-sharing computer. In laboratory it is surrounded by infrared interferometer and is normally used to make Fourier transforms of interferograms.
—FIG. 1

Physical Review from 1959 onto the IBM 7094 operated by Project MAC.³ Project MAC uses a compatible time-sharing system that is available by standard telephone connections through the Bell Telephone TWX or Western Union TELEX teletype machine, and consoles are, therefore, available in laboratories and offices. The teletype machine used in this bibliographical search was obtained primarily for laboratory use, and is shown in figure 1 surrounded by a far-infrared Michelson interferometer for plasma studies. The remote console is normally used to take the Fourier transform of an interferogram and deliver a standard spectrum.

The TIP program contains title, author, reference and entire bibliography of every article covered by the computer. The mode of using this information in preparing the revision of *Basic Data* serves as an illustration of the whole bibliographic search technique. *Basic Data* has chapters titled "Elastic Collision Cross Sections," "Charge-Transfer Cross Sections," "Diffusion and Mobility," "Electron Attachment," etc.

Entrance into the TIP system is made by choosing a key word or words most likely to be contained in the title of an article in a specific subject. Figure 2 illustrates the teletype output from asking the computer to search for all articles in which the words *ion collision* appear. Time is the most important commodity in computer use, and rather than search all the library literature, a file entitled BASIC DATA has been written. It requests a search of certain journals not including theoretical journals, *The Physical Review B*, solid state and nuclear physics journals, and the like.

The machine goes through the journals, printing out the numbers of the volumes searched, and if a title is found, it prints author, title and reference, as shown in figure 2.

Having entered the program with a title search, one may now browse by asking the machine to search out all literature that shares at least one bibliographical reference with any one paper of interest. This procedure has been found from experience to be the most efficient way to find all the rest of the relevant literature independent of whether or not a properly descriptive title was chosen by the author for his particular paper. This procedure is illustrated in figure 3 where we have asked the computer to search the BASIC DATA literature again and find every article that shares at least one bibliographical reference with the paper found by the title "Additional Collision Cross Sections for Helium Especially in the Ionized Continuum." Although it is not illustrated in this figure, we could have asked the machine to type out the particular bibliographical reference or references that provide the link among various articles sharing bibliographies.

Another characteristic of this kind of bibliographic search is also shown in figure 3. When a search is being made for a particular subject, one does not need to search through the entire library because it is characteristic of physicists to use particular journals for particular subjects. After one has made a subject title run, the grouping of these subjects in particular journals is evident, and one need only search a limited number of journals already pinpointed by the title search to do a shared bibliography run for browsing in this subject. One

READ BASIC DATA
FIND TITLE ION COLLISION
OUTPUT PRINT AUTHOR TITLE IDENTIFICATION
GO

APPLIED PHYSICS LETTERS
VOLUME 4
VOLUME 5
VOLUME 6
VOLUME 7

CANADIAN JOURNAL OF PHYSICS
VOLUME 41
J055 1041 P0316
ON THE APPLICATION OF ASTON RANGE AND APPARENT POTENTIAL
MEASUREMENTS TO ION COLLISION STUDIES
MCCORMACK WILLIAM
KERRIN LARKIN

VOLUME 42
J055 1042 P0096
COLLISIONS OF LO-D-LIVED EXCITED IONS OF OXYGEN AND NITROGEN
MCCORMACK WILLIAM
KERRIN LARKIN

VOLUME 43
NORVEGIAN PHYSICAL ACTA
VOLUME 36
VOLUME 37
VOLUME 38

INDIAN JOURNAL OF PHYSICS
VOLUME 38
VOLUME 39

TITLE-SEARCH OUTPUT. Computer was asked to search memory that contained titles, authors and bibliographies from 26 journals for titles in which *ion collision* appears. J055 is code for *Can. J. Phys.* —FIG. 2

can always check whether articles are being lost by this technique, and experience shows that members of the physics community who publish in the 25 most used physics journals do not commonly contribute to atypical journals.

Avoiding duplication

Bibliographies have a great deal of overlap, of course, and when we actually use the program routinely, the machine is asked to search for all the shared bibliographies on a particular subject at the same time so that this duplication does not print out on the teletype output more than once. This not only is a tremendous saving of typing and computer time but also gives the reader an uncluttered record of references in that particular area.

Since entire bibliographies for all articles are contained in the program, we can always ask the computer to print out entire bibliographies of any particular article in which we are interested. These articles come out in a code form illustrated in figure 4. One needs to transform the journal references from numbers to names by a listing that in the present program contains about a thousand journals. With the few illustrations given in this report, one can see what some of these journal numbers are. For example, in figure 2, *Canadian Journal of Physics* is obviously J055 and the *Journal of Chemical Physics* J012, and in figure 3, *The Physical Review* is J001. Since the search

SEARCH JOURNAL OF CHEMICAL PHYSICS
PHYSICAL REVIEW
PHYSICAL REVIEW LETTERS
FIND SHARED BIBLIOGRAPHY JOURNAL OF CHEMICAL PHYSICS AND
OUTPUT PRINT AUTHOR TITLE IDENTIFICATION
GO

JOURNAL OF CHEMICAL PHYSICS
VOLUME 38
VOLUME 39
VOLUME 40
J012 1040 P1271
HIGHER-RESOLUTION STUDY OF THE ELECTRON-IMPACT SPECTRUM OF
HELIUM
SKERDELE AURAK M.
LASSETTRE EDWIN M.

VOLUME 41
VOLUME 42
J012 1042 P0430
GENERALIZED OSCILLATOR STRENGTHS AND ELECTRONIC COLLISION CROSS
SECTIONS FOR NITROGEN AT EXCITATION ENERGIES ABOVE 10 eV
SILVERMAN S. M.
LASSETTRE EDWIN M.

J012 1042 P0430
DETERMINATION OF MOLECULAR EXCITATION POTENTIALS BY ELECTRON
IMPACT: AN EXAMPLE IN THE H₂ SPECTRUM
LASSETTRE EDWIN M.
CLARK FREDERIC M.
MAYR VINCENT G.
SKERDELE AURAK M.

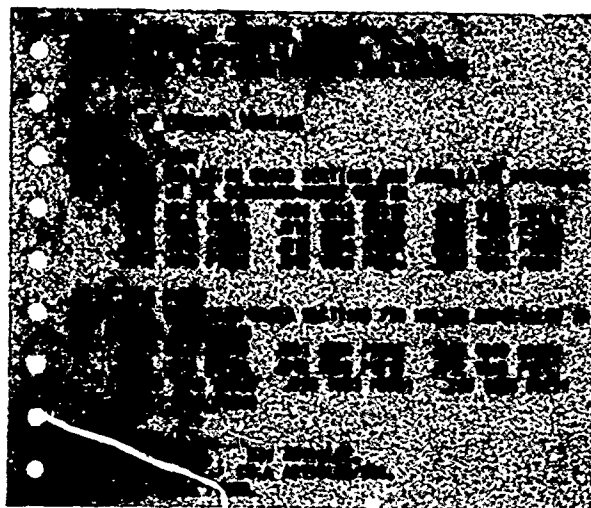
VOLUME 43

BIBLIOGRAPHIC-SEARCH OUTPUT in which computer repeats literature search and finds articles that share bibliographical references with those already selected by title. Method finds most relevant literature. —FIG. 3

illustrated in figure 4 asked for the article by author, an extra article by these authors not containing *ion collision* in the title showed up in the same volume of the *Journal of Chemical Physics*. The illustration was chosen to show an author search as distinct from a title search.

Two kinds of time are to be considered in using a computer, computer processing time and real-time printout. Actual computer time is important for two reasons. With a busy computer, an individual programmer's time is rationed (at MIT one has a monthly allotment) and you want to use your quota with maximal efficiency. Also, computer time costs money so that you are also circumscribed by budget. The principal incentive for developing a special file is to save computer time. The amount saved is, of course, completely dependent on the inclusiveness of the file. With the present storage size of the TIP library program, it takes 273 sec to scan the whole program even if nothing is found (a test was run searching by title for PHYSICS TODAY), whereas to search the entire BASIC DATA file takes only 130 sec.

With a remote console such as illustrated in figure 1 printout time does not cost computer time, since the printout memory is in a small auxiliary computer (IBM 7750) that feeds through a remote console, but it certainly can take operator time. For example, deliberately choosing a time-consuming case, we asked the computer to print out all titles in the BASIC DATA program that con-



BIBLIOGRAPHY PRINTOUT. Computer is asked to list entire bibliographies of selected articles in code. Here, for example, J001 stands for *The Physical Review*. —FIG. 4

tained the word *scatter*. This search, which found 835 titles, took 236 sec of computer processing time but 2 hours and 26 min of real time. During this search, the computer was running at a time of light loading so that it did not have to wait in the compatible time-sharing sequence, and running time was actual typing time.

Saving waste time

A subroutine can be further refined to cut down printout time by asking the program to discriminate against the characteristic words in the title that clearly indicate the paper to be unwanted. Such a TITLE IGNORE list has been developed for the BASIC DATA file, and as an illustration of the type of words that clearly characterize papers outside the realm of gaseous electronics, figure 5 is a copy of the teletype output of the TITLE IGNORE list. Using this kind of title discrimination, the BASIC DATA program searching for the word *scatter* found 496 titles that took about the same computer time but an elapsed time of 1 hour and 16 min.

This last illustration demonstrates one of the drawbacks of using a remote console for obtaining a real-time printout: it is not very fast. If console time is not important—for example, if you start a search and go off and teach a class while it is typing away—long searches are quite satisfactory.

On the other hand our experience shows us that if there is a pressure of use on the console, it is better to use the compatible time-sharing console for the search process and then direct that the output be stored and printed out by the offline printers at the computer site. If this is done, using the last example, the console would be tied up for only about 3 min, instead of 1 hour and 16 min.

A problem that arises when we search in depth in a particular area is that there may be material that, although not of general interest for a public library file, is nevertheless an important source of information in a particular narrow field. This difficulty occurs in gaseous electronics because of the established procedure of publishing in the proceedings of the International Conferences on Ionization Phenomena in Gases. Not only is time valuable, but so also is the number of disk tracks available for archival memory. In the process of the present investigation it has seemed best to put special proceedings papers into a private file that is not part of the main library but which TIP can process and which can be included in the range of literature searched by BASIC DATA. This procedure saves tracks in the total TIP library program. It is actually a simple thing to do, and whenever the BASIC DATA file is called for, not only will the computer search the journals that are common to the total library program, but it will also search the proceedings of the Fifth International Conference on Ionization Phenomena in Gases held in Munich in 1961, the Sixth Conference held in Paris in 1963 and, when it is published, the Seventh Conference held in Belgrade in 1965. It should be pointed out that, in general, conference proceedings do not contain material that has archival value, but in the area of ionization phenomena it has become customary not to overlap the published proceedings papers by

```

TITLE IGNORE      04/12 1966.3
0001  AND NOT T. NUCLE. NOT +NEV NOT NEUTRON NOT RADIATION NOT CRYSTAL
0002  AND NOT T. ISOTOPE NOT +NEV NOT G-RAV NOT BECKE NOT BORN NOT DORN
0003  AND NOT T. SILICON NOT +WOL NOT AROMATIC NOT PROTIN NOT +WOL
0004  AND NOT T. DYE NOT +WOL NOT LIQUID NOT SEMICONDUCT
0005  AND NOT T. METAL NOT +ADHESIVE NOT PRESTON NOT +NEV NOT PI-B
0006  AND NOT T. FILM NOT P-TYPE NOT NICKEL NOT DIELECTRIC NOT HARTREE-FUCK
0007  AND NOT T. P-P NOT COULOMB NOT X-RAY NOT SOLID NOT PHASE
0008  1.003-566

```

TITLE-IGNORE LIST shows titles not searched by computer because it was instructed to discriminate against titles in which certain words indicate that the article probably is not of interest to the data compilers despite other words indicating that it would be. —FIG. 5

publishing elsewhere in various physics journals.

Because in every field there are a certain number of classic papers to which a large number of people refer, another method of browsing through the literature is very simple by the computer technique. The TIP program has been written so that a reference to be cited need not itself be available in the program by title, author, reference and bibliography, but only that it appear in the published references of the articles in the program. Every article that cites the classic paper can be found and brought out on the teletype output. In the particular field of collisions of charged particles, R. B. Brode's paper⁴ in the 1933 *Reviews of Modern Physics* is one of the most referred-to data compilations, and as an illustration of this browsing technique, figure 6 illustrates a search through *The Physical Review* for papers that cited Brode's classic review.

The immediate product of this experiment will be a technical report supplement to *Basic Data of Plasma Physics*. All new data that have been found since the issuance of the book will be reproduced in graphical form as they were in the original book. The references, however, instead of appearing in standard American Institute of Physics form will be in easy computer input form but still readable without a conversion list. Thus, for example, the *Journal of Chemical Physics* will not be listed as an uninformative J012, but rather as the recognizable JCHEM. Also, since the volume number and the date are redundant, the computer is programmed only by volume number. On the other hand the date has significance to the reader; so this is included in the reference form. Standard AIP form for a reference would be

G. J. Schulz, Phys. Rev. A135, 988 (1964)

and the computer would only require

PHYREV V00135 P0988

to which the technical report adds

G. J. Schulz (1964).

As an aside it might be worth pointing out that the AIP practice of calling two different journals by the same name and differentiating them by A's and B's in the pagination (with more proliferation to come) causes such serious trouble with the computer program that TIP gives them different names: PHYREV and PHYREB.

Obviously not all material in the *Basic Data* book has been superseded by work in the last few years, and to provide complete coverage for future searching of the BASIC DATA file by the citation



CLASSIC-PAPER BROWSING turns up articles that refer to well known standard references even when standard references are not stored in the memory for themselves alone. Example shown is for 1933 Brode paper. —FIG. 6

method, a subject matter list is provided of all references not superseded by data contained in the technical report supplement. This list is given in recognizable computer language so that machine citations can be made to both new and old data.

The citation technique provides the method for keeping a data compilation up to date after the manuscript has actually been published. On the assumption that any new work will cite in its bibliography data in the published compilation, providing only that the library program is kept up to date, one can ask it at any time to provide details on new papers that cite the latest data presented in the published compilation. Thus in the revision of *Basic Data of Plasma Physics* the material is created in an open-ended form so that anyone with access to the computer program can search the literature for material that will appear after this report is printed. This feature of the bibliographic search by computers provides a new dimension to the published literature in book form which, without this open-ended feature, is out of date quite generally long before the actual publication date. □

References

1. S. C. Brown, *Basic Data of Plasma Physics* (MIT Press, Cambridge, Mass., 1959).
2. M. Kessler, *PHYSICS TODAY* 18, no. 3, 28 (1965).
3. R. M. Fano in *Proceedings of the Symposium on Computer Augmentation of Human Reasoning*; IEEE Spectrum 2, 56 (1965).
4. R. B. Brode, *Rev. Mod. Phys.* 5, 257 (1933).

BIBLIOGRAPHIC COUPLING EXTENDED IN TIME: TEN CASE HISTORIES*

M. M. KESSLER

Massachusetts Institute of Technology, Cambridge, Mass.

Summary—The methods of bibliographic coupling were applied to 8186 papers in thirty-five volumes of the *Physical Review* (Vol. 77, 1950 to Vol. 111, 1958). The results are reported in the form of ten case histories. Each case was chosen to illustrate a problem in information retrieval.

PREVIOUS reports in this series [1, 2] defined the unit of bibliographic coupling as 'one item of reference used by two papers'. A criterion of coupling between papers was also defined: 'a number of papers constitute a related group, G_A , if each member of the group has at least one coupling unit to a given test paper, P_o . The coupling strength between P_o and any member of G_A is measured by the number of coupling units (n) between them.'

The first report [1] demonstrated the existence of the phenomenon of bibliographic coupling. In the second report [2] the criterion was tested on Vol. 97 of the *Physical Review*, (265 articles). The present report extends the study to a much larger population of papers.

The purpose of our work is to establish a factual background that will guide the design of an experimental science communication system. Such a system will have several modes of operation in order to satisfy the complex needs of scientific workers. An example of a common operational mode might be the following. A working scientist is in the habit of reading several journals in his field. If he is a physicist, he will read the *Physical Review* and perhaps two or three other journals. As he reads these sources he will find a paper that satisfies his needs. His demand upon a communication system might be to, 'search the rest of the literature and find others like it'. This simple request, if it is to be fulfilled through automatic processing, implies the acceptance of a 'search technique', a 'criterion of likeness' and some understanding of what we shall mean by 'the rest of the literature'.

We are satisfied that the technique of bibliographic coupling will serve the needs of model building. We are also satisfied that, initially at least, we may think of the 'criterion of bibliographic coupling' as a working 'criterion of likeness'. There remains the matter of what is to be understood by 'the rest of the literature'. What is the search field?

When a scientist considers searching the literature for relevant papers he looks at a number of journals over some time span. We may think of the search field as a two-dimensional field with points along the vertical axis representing journals, and the horizontal axis representing time. It was shown in [3] that such a field is open along the journal axis. The time axis is also essentially open. Clearly, in order to have a manageable search field one must put bounds on both axes. These bounds must be adjustable to suit the needs of the user. We believe that for our purposes the concept of the 'family matrix' as developed in the fourth report of this series [4] provides a method of setting bounds on the journal axis.

* Supported by a grant from the National Science Foundation to the M.I.T. Libraries.

There remains the question of time. Given a paper and the request to find 'others like it', how far do we search backwards and forwards in time? We cannot decide on an arbitrary number of years that will be right for all subjects and all purposes. Much to be preferred is a method of search that adjusts the time span to the particular needs of the subject and motive of search.

This report illustrates bibliographic coupling and group formation over a large span of time. The search field consisted of 8,186 papers published in thirty-five volumes of the *Physical Review* (Vol. 77, 1950 to Vol. 111, 1958). Ten test papers were chosen as P_o 's. The test papers were picked to illustrate some specific type of problem. The results are given in the form of case histories.

SUMMARY OF CASES

In each case one paper is used as P_o , the test paper. The literature in store is searched automatically for others that match P_o according to the defined criterion of coupling. Each case is discussed as an example of a general problem in retrieval.

Case 1:

P_o = (Vol. 97, page 1059 of the *Physical Review*. We shall use the short notation, 97×1059 .)
'Theory of polarization of nucleons scattered elastically by nuclei'. S. Fernbach, W. Heckrotte and J. V. Lepore.

This is a typical paper in a very active and popular field of research. The first processing yields a list of hundreds of papers. The results show how the list converges with time, both past and future. This is also an example of a case that requires higher order processing to *contract* the list to useful size.

Case 2:

P_o = (97×212) 'New formulation of general theory of relativity'. G. Szekeres.

This P_o is an example of a classical field that is not too active at this time and which produces relatively few papers in the literature. Second and third order processing are used to *expand* the list to useful size.

Case 3:

P_o = (97×201) 'Direct detection of soft radiation above 50 km in the auroral zone'. L. H. Meredith, M. B. Gottlieb and J. A. Van Allen.

This P_o illustrates the case where a widely used technique is applied to a very limited locale. The list needs advanced processing in order to produce expansion and *differentiation* of its various aspects.

Case 4:

P_o = (97×66) 'Effect of strong electrostatic fields on the resistance of tungsten wires in high vacua'. W. J. Deshotels and A. H. Weber.

This P_o illustrates the case of a paper somewhat peripheral to the mainstream of physics. The results indicate the tendency of our processing scheme to push the search in the direction of the underlying physics rather than the specific field of application that the author considered.

Case 5.

$P_o = (97 \times 86)$ 'Age of meteorites by the A^{40} - K^{40} method'. G. J. Wasserburg and R. J. Hayden.

This example is similar to case 4.

Case 6:

$P_o = (97 \times 480)$ 'Scattering of 119-mev pions by deuterium'. N. E. Nagle.

This P_o comes closest to illustrating the average results of retrieval through bibliographic coupling.

Case 7:

$P_o = (77 \times 1)$ 'Local production of mesons at 11,300 ft'. O. Piccioni.

The P_o was chosen because it was serially first in our list of 8186 articles, thus offering the maximum search span (thirty-five volumes) into the future. It illustrates convergence of the list in future time.

Case 8:

$P_o = (112 \times 2135)$ 'Charge properties of the K-meson and hyperon decay interaction'. R. F. Sawyer.

This P_o was chosen because it was serially last in our list of 8186 articles, thus offering the maximum search span into the past. It illustrates convergence with respect to past time.

Case 9:

$P_o = (Acta Physica Sinica, 1957, March, p. 101)$ Translation of Chinese title: 'Energy equations of some nuclei in the vicinity of Pb^{208} '.

This P_o was taken from *Acta Physica Sinica*. Written entirely in Chinese, it illustrates the independence of bibliographic coupling from problems of words and language. This case produced a large list that was treated the same as cases 1 and 6.

Case 10:

$P_o = (Acta Physica Sinica, 1957 March, p. 142)$ Translation of Chinese title: 'A study of the acoustic internal friction of iron, vibrating transversely in a steady magnetic field (between) piezo-electric crystal plates'.

This P_o was also taken from *Acta Physica Sinica*. It is similar to cases 4 and 5.

Case 1:

$P_o = (97 \times 1059)$ 'Theory of polarization of nucleons scattered elastically by nuclei'. S. Fernbach, W. Heckrotte and J. V. Lepore.

This article was picked because in the original study on bibliographic coupling within Vol. 97 it was found to have a very rich association with other papers [2]. It represents a popular field in physics. A search of the thirty-five volumes of the *Physical Review* with this P_o resulted in a group of 322 papers. The distribution of the papers among the thirty-five volumes is plotted in Fig. 1.

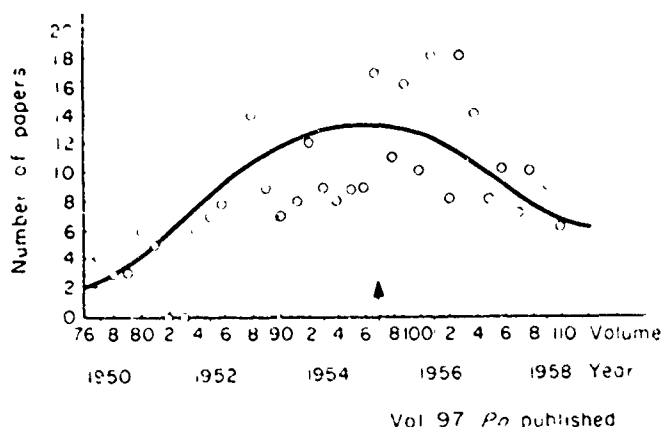


FIG. 1. Time distribution of papers retrieved by $P_o = 97 \times 1059$.

We note from Fig. 1 that the papers in the group occur in volumes fairly symmetrically distributed to either side of the volume containing P_o (Vol. 97, arrow in Fig. 1). The search span of about four or five years to either side of the test paper was not imposed but automatically adjusted by the search method. It is expected that other P_o 's or other search populations might impose a different time span.

A group of 322 papers is too large for most purposes but the scale of graded coupling strengths allows us to use as many articles in the group as will satisfy the given purpose. For illustration, we give below the list of all those papers coupled to P_o with a strength of four or more. The coupling strength is shown in each case.

- $P_o : 11 : (102 \times 1157)^*$ 'Optical-model potential for nucleons scattered by nuclei'. W. B. Rosenfeld and K. M. Watson.
- $P_o : 11 : 102 \times 1659$ 'Experiments with 315-mev polarized protons. I. Elastic scattering by complex nuclei'. O. Chamberlain, E. Segre, R. D. Tripp, C. Wiegand and T. Ypsilantis.
- $P_o : 9 : 97 \times 1077$ 'Polarization in scattering by complex nuclei'. S. Tamor.
- $P_o : 8 : 100 \times 886$ 'Polarization of protons elastically scattered from nuclei'. R. M. Sternheimer.
- $P_o : X : 97 \times 1314$ 'Polarization of nucleons elastically scattered from nuclei'. R. M. Sternheimer.
- $P_o : 6 : 106 \times 1272$ 'Angular distribution of 424-mev polarized protons elastically scattered from various nuclei'. E. Heiberg.
- $P_o : 5 : 89 \times 508$ 'Disintegration of helium by 90-mev neutrons'. P. E. Tannenwald.
- $P_o : 5 : 96 \times 1654$ 'Possible triple-scattering experiments'. L. Wolfenstein.
- $P_o : 5 : 98 \times 1387$ 'High energy cross sections. I. The size of the nucleus'. R. W. Williams.
- $P_o : 5 : 104 \times 449$ 'Polarization of 220-mev protons elastically scattered from carbon'. W. G. Chestnut, E. M. Hafner and A. Roberts.
- $P_o : 5 : 97 \times 1051$ 'Polarization and amplitudes in nucleon-nucleon scattering'. G. Breit, J. B. Ehrman and M. H. Hull Jr.
- $P_o : 5 : 97 \times 1336$ 'Scattering of fast neutrons and protons by atomic nuclei'. G. Takedo and K. M. Watson.
- $P_o : 4 : 98 \times 147$ 'Scattering of polarized nucleon beams'. R. Oehme.

* This notation means that P_o is coupled through eleven coupling units to the paper in Vol. 102, p. 1157 of the *Physical Review*.

- $P_o : 4 : 102 \times 473$ 'Elastic scattering of 340-mev protons by deuterons'. O. Chamberlain and D. D. Clark.
- $P_o : 4 : 104 \times 764$ 'High energy proton-proton scattering and the levy potential'. M. S. Wertheim, M. H. Hull Jr. and A. M. Saperstein.
- $P_o : 4 : 106 \times 793$ 'Diffuse-surface optical model analysis of elastic scattering of 17- and 31.5-mev protons'. M. A. Melkanoff, J. S. Nodnik, D. S. Saxon and R. D. Woods.
- $P_o : 4 : 107 \times 607$ 'Theory of high energy deuteron scattering'. H. P. Stapp.
- $P_o : 4 : 108 \times 463$ 'Direction of polarization by quasi-elastic scattering of 315-mev protons'. H. Broadus and W. Isbell.

Case 2:

$P_o = (97 \times 212)$ 'New formulation of the general theory of relativity'. G. Szekeres.

This paper was chosen for reasons opposite of those in case 1. It is a highly theoretical treatment of one of the most abstract subjects in physics. The theory of relativity is not a very popular research problem among physicists and not many papers are published in the field. In the original study of Vol. 97 this paper did not group with any other paper in that volume. The search with this P_o through the thirty-five volumes of *Physical Review* resulted in a group of only five papers:

- 89×587 'A quasi-relativistic theory of gravitation'. M. G. Moore.
- 98×793 'Use of the flat-space metric in Einstein's curved universe and the "Swiss-Cheese" model of space'. F. J. Belinfante.
- 104×831 'Cosmic time and field equations of general relativity'. W. Cantor and G. Szekeres.
- 104×1791 'Ether drift and gravitational motion'. G. Szekeres.
- 107×1157 'Space-time structure of a static spherically symmetric scalar field'. O. Bergmann and R. Leipnik.

In order to expand the list, each of these five papers was used as P_o for a new search of the entire literature in store. Twenty papers were retrieved in addition to the original five. The results are shown in the list that follows.*

- 79×986 'On the quantization of Einstein's gravitation field equations'. F. A. E. Pirani and A. Schild.
- 80×81 'Hamiltonian of the general theory of relativity with electromagnetic field'. P. G. Bergmann, R. Penfield, R. Schiller and H. Zatzkis.
- 81×1023 'Conservation laws in the general theory of relativity with electromagnetic field'. H. Zatzkis.
- 83×1018 'Constraints in covariant field theories'. J. Heller and P. G. Bergmann.
- 84×665 'Canonical field theory with spinors'. J. Heller and P. G. Bergmann.
- 84×737 'Hamiltonians without parametrization'. R. H. Penfield.
- 87×116 'Quantum theory of interacting gravitational and spinor fields'. B. S. DeWitt and C. M. DeWitt.

* As long as the phenomenon of bibliographic coupling is new and not familiar to most readers we shall continue to publish the titles and authors of the groups formed by this method. We realize that this practice makes our reports bulky and awkward to read. Nevertheless, we consider these lists to be significant data at this stage.

- 87 × 452 'Quantization of Einstein's gravitational field equations, II'. F. A. E. Pirani, A. Schild and R. Skinner.
- 89 × 4 'Classical and quantum field theories in the Lagrangian formalism'. P. G. Bergmann and R. Schiller.
- 89 × 263 'Strong conservation laws and equations of motion in covariant field theories'. J. N. Goldberg.
- 89 × 400 'Spin and angular momentum in general relativity'. P. G. Bergmann and R. Thomson.
- 89 × 1046 'Relativity precession of the asteroid Icarus'. J. J. Gilvarry.
- 92 × 1567 'Equations of motion in Einstein's new unified field theory'. J. Callaway.
- 95 × 243 'Equations of motion of charged test particles in general relativity'. D. M. Chase.
- 95 × 1334 'Fierz-Pauli theory of particles of spin 3/2'. Suraj N. Gupta.
- 96 × 1683 'Gravitation and electromagnetism'. Suraj N. Gupta.
- 97 × 511 'Geons'. J. A. Wheeler.
- 98 × 531 'Dirac bracket transformations in phase space'. P. G. Bergmann and I. Goldberg.
- 98 × 1118 'Special-relativistic derivation of generally covariant gravitation theory'. R. H. Keaichnan.
- 104 × 11 'Application of an artificial satellite to the measurement of the general relativistic "red shift"'. S. F. Singer.

Case 3:

- $P_o = (97 \times 201)$ 'Direct detection of soft radiation above 50 km in the auroral zone'. L. H. Meredith, M. B. Gottlieb and J. A. Van Allen.

In this paper a widely used technique, detection of soft radiation, is applied to a very specific locale, above 50 km in the auroral zone. In its own volume of publication this paper did not form a group. When all the thirty-five volumes were searched only four were retrieved. Of these, two are essentially by the same group of authors. The four papers are:

- 92 × 428 'The time variation of cosmic-ray heavy nuclei'. V. H. Yngve.
- 95 × 147 'Low momentum end of the spectra of heavy primary cosmic rays'. R. A. Ellis Jr., M. B. Gottlieb and J. A. Van Allen.
- 99 × 198 'Cosmic ray intensity above the atmosphere at high latitudes'. L. H. Meredith, J. A. Van Allen and M. B. Gottlieb.
- 102 × 1399 'High altitude cosmic-ray neutron intensity variations'. R. K. Soberman.

We are of the opinion that automatic processing schemes should form groups of papers larger than what the scientist expects to read. The group as a whole should have a high density of useful material but it should be the user who performs the final act of judgment by choosing some titles and discarding others. Although the subject matter of the P_o in this case is admittedly limited, the retrieval of only four papers out of a search population of some 8,200 is not a satisfactory result. We encountered a similar situation in case 2, while in case 1 possibly too many papers were retrieved in the first processing. We suspect that this variation in 'mesh size' is a general property of all uniquely defined retrieval criteria applied automatically to a large variety of papers. It points out the need for 'higher order processing'. In the present case where the original P_o formed a group of only four papers, we use these four papers as P_o 's for a second search through our working literature. When this was

done 406 papers were retrieved. Table 1 shows by reference only those whose coupling strength is $n \geq 4$. As an illustration of the results we list by title and author papers in this group of $n \geq 5$.

TABLE 1. LIST OF PAPERS GENERATED BY PROCESSING THE GROUP FORMED BY $P_o = 97 \times 201$

$P_o = 97 \times 201$			
92 428	95 147	99 × 198	102 × 1399
7 : 90 × 655	5 : 84 × 791	4 : 84 × 572	5 : 80 × 52
8 : 93 × 544	4 : 86 × 536	4 : 88 × 321	9 : 83 × 1175
4 : 94 × 1029	5 : 90 × 655	4 : 90 × 655	4 : 94 × 426
4 : 94 × 1317	5 : 92 × 428	5 : 91 × 957	4 : 95 × 147
5 : 95 × 147	6 : 93 × 544	6 : 93 × 544	6 : 97 × 1276
6 : 95 × 782	4 : 93 × 551	5 : 93 × 551	7 : 100 × 859
5 : 80 × 52	4 : 95 × 782	4 : 93 × 596	6 : 104 × 1718
4 : 103 × 1820	4 : 97 × 1292	7 : 95 × 147	
6 : 108 × 1327	7 : 99 × 198	6 : 95 × 531	
	4 : 102 × 1399	5 : 96 × 1391	
	7 : 108 × 1327	4 : 100 × 1460	
	4 : 110 × 751	4 : 108 × 148	
		4 : 109 × 1367	
$P_o : 5 : 80 \times 52$	'Determination of the high altitude latitude dependence in cosmic-ray neutron intensity'. W. P. Staker.		
$P_o : 9 : 83 \times 1175$	'Neutrons produced in the atmosphere by the cosmic radiations'. J. A. Simpson.		
$P_o : 5 : 84 \times 791$	'Intensities of heavy cosmic-ray primaries by pulse ionization of chamber measurements'. J. A. Van Allen.		
$P_o : 16 : 90 \times 655$	'Fluctuations and latitude effect of cosmic rays at high altitudes and latitudes'. H. V. Neher, V. Z. Peterson and E. A. Stern.		
$P_o : 5 : 91 \times 957$	'Cosmic-ray albedo'. S. B. Treiman.		
$P_o : 20 : 93 \times 544$	'Cosmic radiation in the trapped orbits of a solar magnetic dipole field'. S. B. Treiman.		
$P_o : 9 : 93 \times 551$	'Solar magnetic moment and diurnal variation in cosmic-ray intensity'. J. Firor, F. Jory and S. B. Treiman.		
$P_o : 6 : 95 \times 531$	'Cosmic radiation at very high altitudes near the geomagnetic equator'. M. A. Pomerantz.		
$P_o : 10 : 95 \times 782$	'Primary heavy nuclei'. T. H. Stix.		
$P_o : 5 : 96 \times 1391$	'Composition of the primary cosmic radiation at $\lambda = 10^\circ \text{N}$ '. G. W. McClure.		
$P_o : 6 : 97 \times 1276$	'High-altitude cosmic-ray neutron density at the geomagnetic pole'. H. A. C. Neuburg, R. K. Soberman, M. J. Swetnick and S. A. Korff.		
$P_o : 7 : 100 \times 859$	'Variation of the position of the cosmic-ray neutron intensity maximum with geomagnetic latitude'. R. K. Soberman, A. Beiser and S. A. Korff.		
$P_o : 6 : 104 \times 1718$	'Correlation of meteorological parameters with cosmic-ray neutron intensities'. J. A. Lockwood and H. E. Yingst.		
$P_o : 5 : 108 \times 1327$	'High-altitude intensities of the medium and heavy cosmic-ray nuclei and of the star-producing component over a 25-hour interval'. R. F. Missert.		

Case 4:

$P_o = (97 \times 66)$ 'Effect of strong electrostatic fields on the resistance of tungsten wires in high vacua'. W. J. Deshotel and A. H. Weber.

This paper, according to the authors' abstract, is a re-investigation with better techniques of an experiment performed in 1921. It is an empirical study in a field not too active among physicists at this time. It is to be expected that a paper of this type will not generate many more like it in the *Physical Review*. In its own volume of publication this paper did not group with any others. When the larger body of literature was searched, thirty-seven papers were retrieved. Of these, twenty-one were published prior to P_o and sixteen following P_o .

- 77 × 246 'Thermionic emission of thin films of alkaline earth oxide deposited by evaporation'. G. E. Moore and H. W. Allison.
- 78 × 158 'Average thermionic constants of polycrystalline tungsten wires'. M. H. Nichols.
- 79 × 964 'Photoelectric emission from F-centres in K1'. L. Apker and E. Taft.
- 80 × 887 'Periodic deviations in the Schottky effect for tantalum'. R. J. Munick, W. B. LaBerge and E. A. Coomes.
- 82 × 887 'Some theorems on the free energies of crystal surfaces'. C. Herring.
- 83 × 1005 'Electrical properties of Fe_2O_3 and Fe_2O_3 containing titanium'. F. J. Morin.
- 88 × 655 'A direct comparison of Kelvin and electron beam methods of contact potential measurement'. P. A. Anderson.
- 89 × 244 'Electron ejection from Mo by He^+ , He^{++} and He_2^+ '. H. D. Hagstrum.
- 89 × 851 'Periodic deviations in the Schottky effect for molybdenum'. E. G. Brock, A. L. Houde and E. A. Coomes.
- 90 × 772 'Surface barrier analysis for the highly refractory metals by means of Schottky deviations'. D. W. Juenker, G. S. Colladay and E. A. Coomes.
- 91 × 534 'Photoemission from silver into AgCl, KBr, NaCl, and new brands of photo-sensitivity in AgCl'. M. A. Gilles.
- 91 × 561 'The photoemission of tungsten in the region of predicted Schottky deviations'. E. E. Buder, J. J. Ruddick and A. H. Weber.
- 92 × 637 'The quenching of mercury resonance radiation (2537A) by nitrogen'. C. G. Matland.
- 92 × 1367 'Periodic deviations in the Schottky effect'. S. C. Miller Jr. and R. H. Good.
- 93 × 1136 'Diffusion cooling of electrons in ionized gases'. M. A. Biondi.
- 93 × 1148 'Electron ejection by slow positive ions incident on flashed and gas-covered metallic surfaces'. J. H. Parker Jr.
- 93 × 1249 'The band fluorescence of mercury vapor'. A. O. McConbrey.
- 94 × 295 'Thermionic and surface properties of tungsten crystals'. G. F. Smith.
- 94 × 309 'Growth and surface properties of tantalum crystals'. M. H. Nichols.
- 94 × 910 'Mobilities of atomic and molecular ions in noble gases'. M. A. Biondi and L. M. Chanir.
- 98 × 889 'Velocity analysis of thermionic emission from single-crystal tungsten'. A. R. Hutson.
- 98 × 1831 'Measurement of absorption coefficients for photonizing radiations in low-pressure gases with a space charge detector'. C. D. Maunsell.

- 99 × 1155 'Surface barrier analysis for metals by means of Schottky deviations'. D. W. Juenker.
- 99 × 1307 'Absorption studies of helium metastable atoms and molecules'. A. V. Phelps.
- 100 × 640 'Schottky effect for SrO films on molybdenum'. G. A. Haas and E. A. Coomes.
- 100 × 1115 'Velocity analysis of thermionic emission from single-crystal tungsten'. G. F. Smith.
- 101 × 1694 'Theoretical interpretation of field emission experiments'. J. T. Lewis.
- 107 × 687 'Photoelectric emission from barium oxide'. H. R. Philipp.
- 107 × 1219 'Mobilities of mercury ions in helium, neon and argon'. L. M. Chanin and Manfred A. Biondi.
- 107 × 1553 'Thermionic emission from a planar tantalum crystal'. H. Shelton.
- 108 × 1 'Theory of secondary electron emission by high speed ions'. E. J. Sternglass.
- 109 × 2005 'Dissociative attachment of electrons in iodine. I. Microwave determination of absolute cross section at 300 K'. M. A. Biondi.
- 111 × 103 'Reflection of very slow electrons'. H. A. Fowler and H. E. Farnsworth.
- 111 × 394 'Model for the surface potential barrier and the periodic deviations in the Schottky effect'. P. H. Cutler and J. J. Gibbons.
- 96 × 336 'Theory of Auger ejection of electrons from metals by ions'. H. D. Hagstrum.
- 106 × 473 'Temperature dependence of ion mobilities in helium, neon and argon'. L. M. Chanin and M. A. Biondi.

In examining this list of retrieved papers one must bear in mind that this is not a game played for its own sake. There must be a purpose or a motive for a scientist to pick this particular P_o and ask that thirty-five volumes of the *Physical Review* be searched for 'others like it'. What could this motive be? Of the thirty-seven papers in the list, fifteen concern Field, Thermionic or Photo-Emission. Seven mention the Schottky Effect, five have Tungsten in the title. The remaining twelve concern electrical properties of crystals, etc. It is interesting to note that the original P_o does not mention the Schottky Effect in its title, the abstract, nor in the text of the paper. One could not expect the Schottky Effect to be the reason why the scientist chose this paper for a search. If the motive for search were Field, Thermionic or Photo-Emission, it is also hard to believe that this paper would be chosen as the search medium. One is left with the possibility that the user was primarily interested in the electrical properties of Tungsten. The five papers that mention Tungsten in their titles were therefore used as P_o for a further search. This new search resulted in sixty-three papers. If we consider only those papers for which $n \geq 3$ we end up with twenty-seven articles. Of these twenty-seven, all but four occurred in the first list of thirty-seven. The four new papers, listed below, all concern field emission:

- $P_n : 4 : 89 \times 799$ 'Field emission: large current densities, space charge, and the vacuum arc'. W. P. Dyke and J. K. Trolan.
- $P_o : 3 : 91 \times 1043$ 'Field emission initiated vacuum arc. I. Experiments on arc initiation'. W. P. Dyke, J. K. Trolan, E. E. Martin and J. P. Barbour.
- $P_o : 3 : 100 \times 1619$ 'Field emission microscopy of an allotropic transformation: α - β titanium'. E. G. Brock.
- $P_o : 3 : 104 \times 660$ 'Photoelectric work functioning analysis of emission in an accelerating field'. P. E. Carroll.

If we include papers for which $n = 2$, the following additional articles are retrieved:

- 91 × 1447 'The linear combination in β decay'. D. C. Peaselee.

- 92 × 45 'Space-charge effects in field emission'. J. P. Barbour, W. W. Dolan, J. K. Trolan, E. E. Martin and W. P. Dyke.
- 97 × 1579 'Field emission from rhenium: emission pattern corresponding to hexagonal crystal structure'. G. Barnes.
- 101 × 1460 'Comparison of the characteristic energy losses of electrons with the fine structure of the X-ray absorption spectra'. L. B. Leder, H. Mendlowitz and L. Martin.
- 102 × 340 'Auger electron emission in the energy spectra of secondary electrons from Mo and W'. G. A. Harrower.
- 106 × 921 'Resonance potentials in thin films of potassium chloride'. H. Jacobs, I. N. Greenberg, L. Goble and A. Ramsa.

One could continue the search by examining also the papers for which $n = 1$. We shall not go through this exercise because the data already shown is suffice to illustrate the complexity of the relation between a scientist and his literature. This is frequently, but erroneously referred to as 'man-machine interaction'. More precisely it is an interaction between *man and the recorded literature*. In our case there is a constant tendency of the literature to push the user in the direction of articles concerned with emission phenomena and the Schottky effect.* This is the result of the interaction of a man, who chose this P_0 as a test paper, and the literature recorded in the *Physical Review*. Had he chosen the same P_0 but a different literature to interact with, the results might have been quite different.

Case 5:

$P = (97 \times 86)$ 'Age of meteorites by the A^{40} - K^{40} method'. G. J. Wasserburg and R. J. Hayden.

This P_0 is an example of the application of an active technique (nuclear) to a very specialized problem (meteorites). A consideration of the technique alone would, we suspect, have generated a very large list of related papers. Its application to meteorites, however, limited the scope of relevance with the result that only twelve papers out of 8,186 were found relevant by our criterion. Of the twelve, five appeared before P_0 and seven after. The list follows:

- 79 × 450 'A redetermination of the relative abundances of the isotopes of neon, krypton, rubidium, xenon and mercury'. A. O. Nier.
- 79 × 490 'Decay constants of K^{40} '. G. A. Sawyer and M. L. Wiedenbeck.
- 79 × 940 'Scintillation studies on potassium iodide'. B. Smaller, J. May and M. Friedman.
- 90 × 857 'Argon 38 in pitchblende minerals and nuclear processes in nature'. W. H. Fleming and H. G. Thode.
- 93 × 749 'Elastic scattering of alpha particles by neon'. E. Goldberg, W. Haberli, A. I. Galonsky and R. A. Douglas.
- 99 × 771 ' K^{40} branching ratio'. A. McNair, R. N. Glover and H. W. Wilson.
- 99 × 1747 'Densities and imperfections of single crystals'. A. Samkula, J. Kalnays and V. Sils.
- 101 × 1786 'Search for possible naturally occurring isotopes of low abundance'. F. A. White, T. L. Collins and F. N. Rourke.

* The two, of course, are intimately related.

- 103 × 987 'Decay constants of K^{40} as determined by the radiogenic argonne content of potassium minerals'. G. W. Weatherill, G. J. Wasserburg, L. T. Aldrich, G. R. Tilton and R. J. Hayden.
- 105 × 765 'Crucial experiment concerning the origin of meteorites'. S. F. Singer.
- 107 × 540 'Radiation age of a meteorite from cosmic-ray-produced He^3 and H^4 '. F. Bergmann, J. Geiss and D. C. Hess.
- 107 × 1695 'Depth variation of tritium and argon-37 produced by high-energy protons in iron'. E. L. Fireman and J. Zahringer.

In our approach to the problems of retrieval and communication we do not favour extensive word and concept analysis. The reason is well-illustrated in this very short list of twelve papers. If we attempt to analyse the various words and concepts in the list, we are impressed with the great varieties of expression, shades of meaning and multiplicity of colour and insinuation that may be found in just a few words. Such richness of expression adds significantly to the meanings being communicated but hampers perhaps decisively automatic processing based on words and language. The short title of our paper is 'age of meteorites by the A^{40} - K^{40} method'. The two relevant concepts that the author connects in his paper are expressed by the two phrases 'age of meteorites' and the ' A^{40} - K^{40} method'. Of the twelve papers, only one contains the first phrase and one the second. A more drastic breakdown of the two phrases yields: 'age', 'meteorites', ' A^{40} ' and ' K^{40} '. This breakdown does untold violence to the conceptual structure of the papers. Nevertheless, we note that two papers mention meteorites, three mention K^{40} , none mention A^{40} . What are some possible reasons for a scientist wanting to pursue this P_o into the literature? Depending on the needs of each individual scientist, the search could branch in the direction of K^{40} (potassium 40), meteorites or perhaps into the general field of the natural abundance of various radioactivities. In the first case we would use papers 79 × 490, 99 × 771 and 103 × 987 as P_o and continue the search. This yields, in addition to some repetitions, one new paper:

- 91 × 1223 'The branching ratio of potassium 40'. R. D. Russel and H. A. Shilliber.

If the user were interested in meteorites, he would continue the search with papers 105 × 765 and 107 × 540 of the original list. The first of these does not yield any additional information. The second, entitled 'Radiation age of meteorites from cosmic-ray-produced He^3 and H^4 ' retrieves four papers:

- 97 × 1303 'Tritium production by 2.2 bev protons on iron and its relation to cosmic radiation'. E. L. Fireman.
- 101 × 1557 'Tritium production by high-energy protons'. L. S. Currie, W. F. Libby and R. L. Wolfgang.
- 107 × 1695 'Depth variation of tritium and argon-37 produced by high energy protons in iron'. E. L. Fireman and J. Zahringer.
- 110 × 175 'Production of He^6 by high-energy protons'. F. S. Rowland and R. L. Wolfgang.

These papers are more concerned with tritium production than with meteorites. If the interest of the inquirer is directed toward natural abundance of radioactive materials he would continue the search with papers 79 × 450, 90 × 857 and 101 × 1786. These yield only two additional papers:

- 77 × 789 'A redetermination of the relative abundances of the isotopes of carbon, nitrogen, oxygen, argon and potassium'. A. O. Nier.

88 × 642 'The specific alpha-activities and half lives of U^{234} , U^{235} and U^{236} '. E. H. Fleming, A. Ghiorso and B. B. Cunningham.

This example, as the previous one, indicates the complexity of following the literature in a subject largely peripheral to the mainstream of physics. Meteorites are not a significant occupation among physicists. Papers on this subject perhaps occur in journals other than the *Physical Review*. If, however, we insist on searching the *Physical Review*, we are constantly nudged by the literature in the direction of the physics of the situation rather than the particular application at hand. Thus in case 4 all attempts to follow the literature in the direction of 'tungsten' brought us eventually to electron emission and the Schottky effect. In the present case, all attempts to follow 'the age of meteorites' bring us back to the physical tools used in the study of the age of meteorites. This indicates to us that the literature of science is more than just a store of factual information but has a logic and dynamics that may suggest interconnections that are seemingly not obvious at first sight. This is particularly well-illustrated in case 4.

Case 6:

$P_0 = (97 \times 480)$ 'Scattering of 119-mev pions by deuterium'. N. E. Nagle.

This P_0 was chosen because in searching its volume of publication it yielded rather pedestrian results. Six papers were grouped, all of minimal coupling strength ($n = 1$):

- 97 × 791 'Scattering of 151-mev positive pions by protons'. R. A. Grandey and A. F. Clark.
- 97 × 902 'Thermal conductivity of indium-thallium alloys at low temperatures'. R. J. Sladek.
- 97 × 1059 'Theory of polarization of nucleons scattered elastically by nuclei'. S. Fernbach, W. Heckrotte and J. V. Lepore.
- 97 × 1071 'Specific ionization by high-energy electrons'. W. Barber.
- 97 × 1336 'Scattering of fast neutrons and protons by atomic nuclei'. G. Takeda and K. M. Watson.
- 97 × 1344 'Approximate reduction of the many-body problem for strongly interacting particles to a problem of self-consistent fields'. K. A. Brueckner and C. A. Levinson.

A search of the thirty-five volumes yielded 183 papers. Their distribution in time is plotted in Fig. 2.

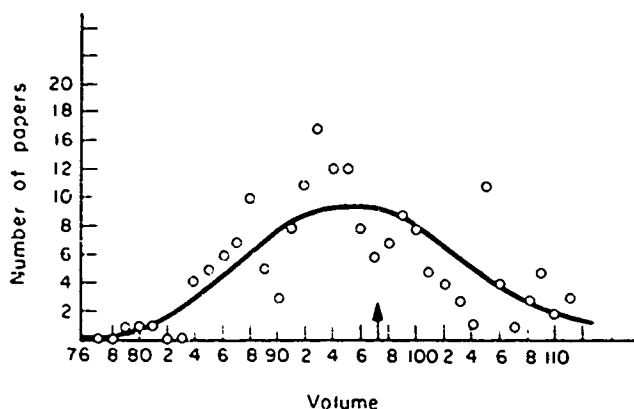


FIG. 2. Time distribution of papers retrieved by $P_0 = 97 \times 480$.

In this group of 183, 142 papers had a coupling strength of one, twenty-nine had $n = 2$. The twelve remaining papers of $n \geq 3$ we list by title and author:

- $P_o : 9 : 105 \times 247$ 'Interactions of 85-mev positive pions with deuterons'. K. C. Rogers and L. M. Lederman.
- $P_o : 9 : 109 \times 1733$ 'Interactions of 38- and 61-mev positive pions in deuterium'. A. M. Sachs, H. Winick and B. A. Wooten.
- $P_o : 6 : 91 \times 135$ 'A diffusion cloud-chamber study of pion interactions in hydrogen and helium'. E. C. Fowler, R. P. Shutt, A. M. Thorndike and W. L. Whittemore.
- $P_o : 5 : 96 \times 1104$ 'Total cross sections for negative and positive pions in hydrogen and deuterium'. J. Ashkin, J. P. Blaser, F. Feiner, J. G. Gorman and M. O. Stern.
- $P_o : 5 : 99 \times 849$ 'Multiple scattering corrections in π^- deuteron scattering'. S. D. Drell and L. Verlet.
- $P_o : 4 : 89 \times 575$ 'Multiple scattering and the many-body problem—applications to photomeson production in complex nuclei'. K. M. Watson.
- $P_o : 4 : 89 \times 834$ 'Multiple scattering corrections to the impulse approximation in the two-body system'. K. A. Brueckner.
- $P_o : 4 : 91 \times 960$ 'The scattering and absorption of π^\pm deuteron scattering'. J. F. Tracy.
- $P_o : 4 : 101 \times 371$ 'Photodisintegration of the deuteron'. F. Zachariasen.
- $P_o : 4 : 105 \times 256$ 'Pion deuteron scattering in the impulse approximation'. R. M. Rockmore.
- $P_o : 3 : 88 \times 621$ 'The impulse approximation and field theoretical calculations'. J. S. Blair and B. Segall.
- $P_o : 3 : 94 \times 1335$ 'Scattering of 217-mev negative pions by hydrogen'. M. Glicksman.

A comparison of this list with the test paper ($P_o = 97 \times 480$) confirms that as a general trend the higher 'n' numbers represent papers of closer relation to the search article. This conclusion is further strengthened by comparing this group of relatively high 'n' numbers with the group of papers retrieved from Vol. 97 where 'n' was one in all cases.

We also call attention to Fig. 2 where convergence is again illustrated.

Case 7:

$P_o = (77 \times 1)$ 'Local production of mesons at 11,300 ft'. O. Piccioni.

This P_o is serially the first in the search population of 8,186 papers. It thus provides the largest search span into the future. It was chosen to illustrate the phenomenon of conver-

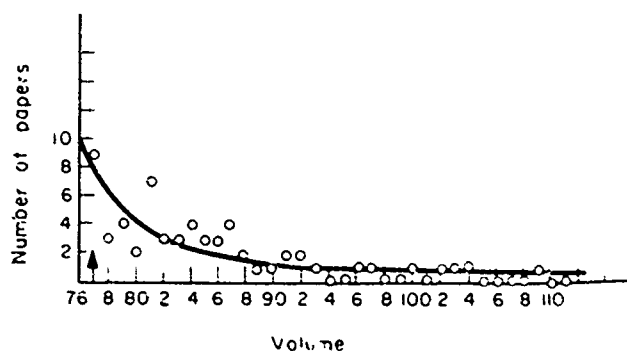


FIG. 3. Time distribution of papers retrieved by $P_o = 77 \times 1$.

gence on the 'future' side of the graph. The search yielded sixty-one papers all, of course, published after P_o . The distribution among volumes is shown in Fig. 3.

Fig. 3 resembles and extends the right-hand sides of the illustrations in Figs. 1 and 2. Of the sixty-one papers collected in this group, forty-four had a coupling strength of one, the remaining seventeen of strength $n \geq 2$, are listed below by title and author.

$P_o : 6 : 77 \times 102$	'On the disintegration products of the 2.2μ s. meson'. E. P. Hincks and B. Pontecorvo.
$P_o : 6 : 79 \times 749$	'Experiments on cosmic-ray mesons and protons at several altitudes and latitudes'. M. Conversi.
$P_o : 4 : 80 \times 619$	'Latitude and altitude dependence of the local hard showers of cosmic rays'. T. G. Walsh and O. Piccioni.
$P_o : 3 : 83 \times 1085$	'Measurement of the positive π -meson lifetime'. C. E. Wiegand.
$P_o : 3 : 84 \times 684$	'Experimental search for the beta-decay of π^+ mesons'. H. L. Friedman and F. Rainwater.
$P_o : 3 : 85 \times 161$	'Production cross sections for π^+ and π^- -mesons by 340-mev protons on carbon and lead at 90° to the beam'. C. Richman, M. Weissbluth and H. A. Wilcox.
$P_o : 3 : 89 \times 983$	'The gamma rays from negative μ -meson capture in lead'. G. G. Harris and T. J. B. Shanley.
$P_o : 2 : 77 \times 180$	'Low energy mesons in the atmosphere'. M. Sands.
$P_o : 2 : 77 \times 342$	'A cloud-chamber study of cosmic-ray nuclear interactions at 3260 m elevation'. W. W. Brown and A. S. McKay.
$P_o : 2 : 77 \times 686$	'A study of penetrating showers at 3260 m'. W. D. Walker.
$P_o : 2 : 81 \times 565$	'The gamma ray spectrum resulting from capture of negative π -mesons in hydrogen and deuterium'. W. K. H. Panofsky, R. L. Aamodt and J. Hadley.
$P_o : 2 : 82 \times 359$	'Nuclear interaction in gold of secondary particles emitted in penetrating showers'. A. J. Hartzler.
$P_o : 2 : 85 \times 891$	'The emission of slow positive and negative mesons from nuclear disruptions produced by cosmic radiation'. H. Yagoda.
$P_o : 2 : 87 \times 942$	'Capture of μ -mesons in heavy elements'. J. W. Keuffel, F. B. Harrison, T. N. K. Godfrey and G. T. Reynolds.
$P_o : 2 : 91 \times 971$	'The mean life of negative μ -mesons stopped in iron'. A. H. Benade.
$P_o : 2 : 92 \times 134$	'The variation of the ratio of positive to negative cosmic-ray μ -mesons with momentum and altitude'. H. A. Morewitz and M. H. Shamos.
$P_o : 2 : 79 \times 952$	'On nuclear evaporation in cosmic rays and the absorption of the nucleonic component, II'. G. Bernardini, G. Cortini and A. Manfredini.

Case 8:

$P_o = (112 \times 2135)$ 'Charge properties of the K-meson and hyperon decay interaction'. R. F. Sawyer.

This P_o is the last in our data population. It thus provides the largest search span into the past. The distribution among volumes is shown in Fig. 4.

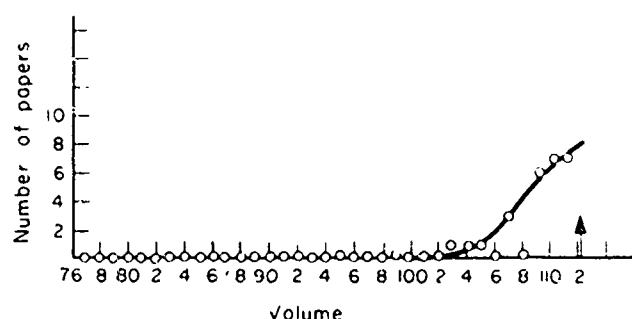


FIG. 4. Time distribution of papers retrieved by $P_o = 112 \times 2135$.

The article was chosen to illustrate convergence on the 'past' side of the graph. The search yielded only twenty-six papers all, of course, published prior to P_o . The list is small enough to be given in its entirety:

- 103 x 1111 'Charge conjugation and the τ^0 meson'. G. A. Snow.
- 104 x 1164 'Dynamical theory of K mesons'. J. Schwinger.
- 105 x 258 'Decay of charged K mesons'. R. Arnowitt and W. B. Teutsch.
- 107 x 573 'Isotopic spin selection rule $|\Delta I| = \frac{1}{2}$ for the decay of strange particles'. M. Kawaguchi.
- 107 x 1396 'Possibility of hyperfragment formation in K^- -d reactions'. A. Pais and S. B. Treiman.
- 107 x 1714 'Pion distribution to hyperon-nucleon forces'. D. B. Lichtenberg and M. H. Ross.
- 109 x 989 'Furry's theorem of very strong interactions'. R. E. Pugh.
- 109 x 1755 'Possible determination of the spin of Δ^0 from its large decay angular asymmetry'. T. D. Lee and C. N. Yang.
- 109 x 1759 'Polarization effects in Σ^- capture'. A. Pais and S. B. Treiman.
- 109 x 2160 'Possible model for strong interactions'. S. Barshay.
- 109 x 2163 'K-meson contribution to forces between baryons'. D. R. Lichtenberg and M. Ross.
- 109 x 2177 'Calculation of the anomalous magnetic moments of the Δ^0 and the Σ^0 , Σ^+ , Σ^- hyperons'. M. Nauenberg.
- 110 x 569 'K-meson dispersion relations. II. Applications'. P. T. Matthews and A. Salam.
- 110 x 574 'Symmetries of the strong interactions'. A. Pais.
- 110 x 743 'Scattering of K^+ mesons by nucleons'. S. Barshay.
- 110 x 1200 'General covariance and elementary particles'. R. Finkelstein.
- 111 x 337 'Pion decay and possible nonlocal effects in the theory of fermi interactions'. A. Sirlin.
- 111 x 354 'Form factors in β decay and μ capture'. M. L. Goldberger and S. B. Treiman.
- 111 x 362 'Test of the nature of the vector interaction in β decay'. M. Gell-Mann.
- 111 x 632 'Production of strange particles by π nucleon and photonucleon interaction near threshold'. R. K. Adair.
- 111 x 957 'Two-component fermion theory'. L. M. Brown.
- 111 x 967 'Hypernuclear binding energies and the Δ -nucleon interaction'. R. H. Dalitz and B. W. Downs.
- 111 x 986 'Theory of internal space'. W. E. Thirring.

111 × 1426 'Possible experimental test of universal fermi interaction'. R. Gatto.

111 × 1433 'Scattering amplitudes for the Thirring model'. F. L. Scarf.

111 × 1691 'Photon decay of hyperons'. R. E. Behrends.

Case 9:

P_o = (*Acta Physica Sinica*, 1957, March, p. 101) Translation of Chinese title: 'Energy equations of some nuclei in the vicinity of Pb^{208} '.

This paper is written entirely in Chinese; a Russian abstract is appended. Although there was no access to the text, the bibliography was in the script of the journals referred to. A glance at the bibliography indicated that this journal fits well into the *Physical Review* family matrix as a receptor journal [4]. It should therefore enter into meaningful group association with other papers in the *Physical Review*. One hundred and ninety papers were retrieved from the thirty-five volumes. This large number is to be expected considering the active nature of the research subject. The distribution of articles among volumes is shown in Fig. 5.

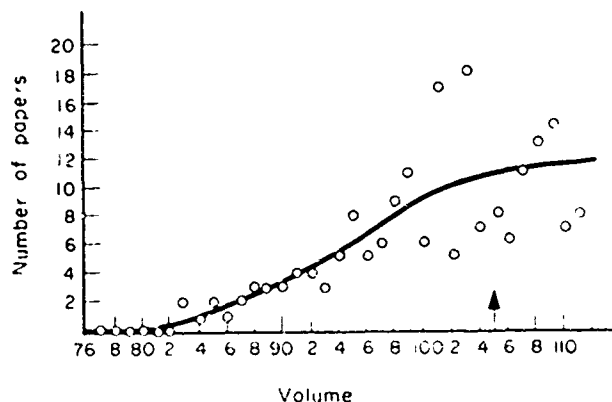


Fig. 5. Time distribution of papers retrieved by P_o from *Acta Physica Sinica*.

Of the 190 papers, 147 coupled to this P_o with $n = 1$, seventeen with $n = 2$, and five with $n = 3$. The remaining twenty papers of $n \geq 4$ are listed below by title and author:

P_o : 6 : 100 × 891 'Theory of neutron reactions with nuclei at low energy'. K. A. Brueckner, R. J. Eden and N. C. Francis.

P_o : 5 : 103 × 172 'Effects of nondegeneracy of nuclear ground state on low energy neutron reactions'. K. A. Brueckner.

P_o : 5 : 103 × 420 'Formal theory of nuclear models'. N. Fukada.

P_o : 5 : 103 × 1353 'Nuclear many-body problem'. H. A. Bethe.

P_o : 5 : 103 × 1558 'Energy level shifts in a large enclosure'. N. Fukada and R. G. Newton.

P_o : 5 : 109 × 447 'Energy levels of polonium-210'. R. W. Hoff and J. M. Hollander.

P_o : 4 : 98 × 1445 'High energy reactions and the evidence for correlations in the nuclear ground state wave function'. K. A. Brueckner, R. J. Eden and N. C. Francis.

P_o : 4 : 99 × 76 'Nuclear energy level fine structure and configuration mixing'. K. A. Brueckner, R. J. Eden and N. C. Francis.

P_o : 4 : 99 × 1418 'Nuclear saturation: a generalized Hartree-Fock method'. R. J. Eden.

P_o : 4 : 102 × 767 'Gamma rays from neutron inelastic scattering'. R. B. Day.

- $P_o : 4 : 103 \times 1008$ 'Nuclear saturation and two body forces: self-consistent solutions and the effects of the exclusion principle'. K. A. Brueckner and W. Wada.
- $P_o : 4 : 104 \times 401$ 'Nuclear energy levels in a velocity-dependent potential'. A. A. Ross, R. D. Lawson and Hans Mark.
- $P_o : 4 : 104 \times 492$ 'Energy of a many-particle system'. W. B. Riesenfeld and K. M. Watson.
- $P_o : 4 : 104 \times 1069$ 'Scattering of 4.4-mev neutrons by aluminum, calcium, chromium and bismuth'. J. B. Weddell.
- $P_o : 4 : 104 \times 1077$ 'Spin-orbit interactions in nuclei'. L. S. Kisslinger.
- $P_o : 4 : 107 \times 1028$ 'Interaction of a nucleon with the nucleus'. E. Vogt and J. Lascoux.
- $P_o : 4 : 108 \times 518$ 'Equation of state of gases and liquids at low temperature'. W. B. Riesenfeld and K. M. Watson.
- $P_o : 4 : 109 \times 1023$ 'Properties of nuclear matter'. K. A. Brueckner and J. L. Gammel.
- $P_o : 4 : 109 \times 1623$ 'Brueckner's theory and the method of superposition of configurations'. R. K. Nesbet.
- $P_o : 4 : 110 \times 431$ 'Theory of finite nuclei'. K. A. Brueckner, J. L. Gammel and H. Weitzner.
- $P_o : 4 : 110 \times 915$ 'Nuclear levels excited by inelastic neutron scattering'. F. I. Boley, E. H. Thorndike and A. T. Moffet.
- $P_o : 3 : 99 \times 792$ 'Independent particle model of the nucleus. I. Interparticle forces and configuration mixing'. C. Levinson and K. W. Ford.
- $P_o : 3 : 100 \times 83$ 'Gamma rays excited by inelastic scattering of neutrons in various elements'. M. A. Rothman, H. S. Hans and C. E. Mandeville.
- $P_o : 3 : 103 \times 489$ 'Applications of scattering theory to quantum statistical mechanics'. K. M. Watson.
- $P_o : 3 : 103 \times 1121$ 'Relation between nucleon density and nuclear potential'. K. A. Brueckner.
- $P_o : 3 : 110 \times 597$ 'Single particle energy and effective mass and the binding energy of many-body systems'. K. A. Brueckner.

This example illustrates that bibliographic coupling is independent of words and language. It further illustrates that the publication of an article in a particular journal is a matter of local choice and convenience. The article in this case could have appeared in the *Physical Review* or any of a number of journals related by the conditions defined for the family matrix [4]. From the point of view of a science information system all the papers occurring in any member of a related family may be processed together and individualized journals formed whose content is dictated by the P_o 's submitted.

Case 10:

$P_o =$ (*Acta Physica Sinica*, 1957, March, p. 142) Translation of Chinese title: 'A study of the acoustic internal friction of iron, vibrating in a steady magnetic field (between) Piezo-electric crystal plates'.

This is an illustration similar to case 4, an empirical treatment of a somewhat peripheral field in modern physics. Not too many papers in the *Physical Review* are expected to group with this P_o . Those found are likely to tend to the theoretical side of the problem and stress the basic science involved. Eleven papers were retrieved:

- 78×420 'Anomalous internal friction associated with the precipitation of copper in cold-worked Al-Cu alloys'. Ting-Sui Ke.

- 80 × 249 'Variation of amplitude-dependent internal friction in single crystals of copper with frequency and temperature'. A. S. Nowick.
- 83 × 709 'Precipitation of carbon and nitrogen in cold-worked alpha-iron'. S. Harper.
- 91 × 816 'The variation of elastic constants of crystalline aluminum with temperature between 63° K and 773° K'. P. M. Sutton.
- 88 × 925 'Anelastic measurements of atomic mobility in substitutional solid solutions'. A. S. Nowick.
- 86 × 898 'Magnetostrictive vibration of prolate spheroids Ni-Fe and Ni-Cu alloys'. J. S. Kouvelites and L. W. McKeethan.
- 94 × 1573 'Moduli and internal friction of magnelite as affected by the low-temperature transformation'. M. E. Fine and Nancy T. Kenny.
- 96 × 296 'Elastic moduli of tin at the superconducting transition'. J. K. Landauer.
- 103 × 294 'Atomic mobility in a Cu-Al alloy after quenching and neutron irradiation'. Chi Yao Li and A. S. Nowick.
- 108 × 690 'Mechanical resonance dispersion in metals at audio frequencies' E. R. Fitzgerald.
- 110 × 1069 'Elastic constants of strontium titanate'. E. Poindexter and A. A. Giardini.

Cases 9 and 10 illustrate a form of processing that is usually neglected. When science communication systems are considered, the stress is on retrieval and the image one has in mind is that of an individual scientist facing a large body of undifferentiated literature. We seldom consider the corollary to this; that is, the individual paper or document facing a large body of undifferentiated readers and, for one reason or another, not getting through. It is fair to say that every serious scientific paper ought to be exposed to the right community of scientists just as every serious scientific worker ought to be exposed to the right literature. In neither case is it a matter of all and everything. A scientific paper need not be read by 'all' scientists potentially concerned just as a working scientist need not read 'everything' in the field. A scientist needs as much literature as will enable him to work efficiently. With equal strength one can say that a paper has to be read by enough scientists in order to be evaluated and integrated into the main stream of the literature.

There are several reasons why a potentially worthy paper may be neglected by a community of scientists. Cases 9 and 10 in this report treated one of the most common of those reasons, the language barrier. These papers were written entirely in Chinese and are essentially not available to the American physicists. It is not necessary that these papers be published and distributed to *all* American physicists. It would suffice to send the translation to those American physicists whose papers group with the Chinese paper. We thus expose the Chinese papers to a readership closest to them in interest and preoccupation. The Chinese work can be evaluated and if it is worthwhile, it will be quoted and integrated into the main stream of American and Western physics. The above considerations apply to literature other than Chinese and to reasons for neglect other than the language barrier.

SUMMARY OF RESULTS

(1) The methods of bibliographic coupling can be applied to a large body of literature to yield meaningful grouping of papers.

(2) The process operates both on the past and future literature, as measured from time of P_0 , the test paper.

(3) The process has the tendency to make the list converge in time. This is equivalent to confining the search to some operational life span of the literature.

(4) Second and third order processing provide an interaction between the reader and the literature. We show examples of such interaction in three cases:

(a) Where the list was too large and we asked for *contraction*.

(b) The list was too small and we asked for *expansion*.

(c) The list showed several branches and we wished to follow one.

(5) The process was applied successfully to Chinese papers where the text was not available for word or content analysis.

REFERENCES

- [1] M. M. KESSLER: An experimental study of bibliographic coupling between technical papers, I.E.E.E Transactions PTGIT, IT-9, 49 (1963).
- [2] M. M. KESSLER: Bibliographic coupling between scientific papers, AM. DOC. 14, 10, 1963.
- [3] M. M. KESSLER and F. E. HEART: Analysis of bibliographic sources in the *Physical Review* (Vol 77, 1950 to Vol. 112, 1958), *M.I.T.*, R-3.
- [4] M. M. KESSLER: Analysis of bibliographic sources in a group of physics-related journals, *M.I.T.*, R-4 (1962).

Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing*

A detailed comparison of how 234 papers in Vol. 112 of *Physical Review* form related groups according to two criteria of relatedness. The criteria are: (1) the Analytic Subject Index as used by the editors of the

Physical Review; and (2) the method of Bibliographic Coupling. The similarities and differences between the groups thus formed are illustrated and discussed.

M. M. KESSLER

*The Libraries, Massachusetts Institute of Technology
Cambridge, Massachusetts*

• Introduction

Previous reports in this series¹⁻³ described and illustrated a new method for separating a large body of technical literature into small related groups. This method was called "Bibliographic Coupling" because it originated in the hypothesis that the bibliography of technical papers is one way by which the author can indicate the intellectual environment within which he operates, and if two papers show similar bibliographies, there is an implied relation between them. The previous reports demonstrated that the phenomenon of bibliographic coupling exists and that the titles so coupled do indeed show a relatedness in subject matter.[†]

The question naturally arises, how does this method of grouping papers compare with the results of certain standard and well-known techniques currently in use? Since the bulk of our experimental material derives from physics literature and, more specifically, from the *Physical Review*, we shall perform the present comparison also on a sample volume of the *Physical Review*.

It is clear that any comparison of the results of two complex processes will reflect the subjective bias and

criteria of the author. It is not just a matter of displaying the author's favoritism or possessive pride. Such crude prejudices can be overcome, and if not they are easily detected by the reader. But the mental bias and the perfectly legitimate intellectual convictions of the author cannot, and perhaps should not, be eliminated. The subjective bias is stamped on the problem and on its solution by the choice of questions that the research is supposed to answer.

The question—How do two processes compare with regard to their ability to form groups of technical papers?—and the question—How do two processes compare with regard to solving the problems of technical communication?—are two different orders of question from the logical point of view. The first question can be answered entirely within the domain of the two processes and their results. Indeed, one may be purely formal and assign arbitrary numbers and letters to the papers and to the subject categories and make the comparison without any reference to the titles and subject fields involved. The result of such a study is apt to be a set of numbers and statistics of considerable precision. But this precision is *not* transferrable to the second order of questions. There we must go out of the domain of the two processes and seek judgment that is relevant to human needs. But such needs reflect complex intellectual habits, tastes and shades of meaning. Do what we will with our statistics, graphs and tables, we must at last face the fact that we are dealing with a field of experience where there is no single criterion that can be applied or definition sharply drawn. A group of papers is more or less

* Supported by a grant from the National Science Foundation to the M.I.T. Libraries and in part by Project MAC, M.I.T. research program sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research.

† The essentials of the method of Bibliographic Coupling are:

- a. A single item of reference used by two papers is called one unit of coupling between them.
- b. A number of papers constitute a related group, G , if each member of the group has at least one coupling unit to a given test paper P .
- c. The coupling strength between P and any member of G is measured by the number of coupling units (n) between them.

what a reader has asked for. A bibliographic list is more or less complete. A given paper fits more or less in a subject field category. The pretension under such circumstances to be rigorously "scientific" and numerically "exact" would only mislead us in our judgment.

Tests and comparisons designed to answer questions of the first order can profitably be performed within the limited environment of the laboratory. Questions of the second order, however, are more suitably tested in a carefully controlled but realistically functional environment that simulates many of the influences and pressures of the real world. It is only to the first type of question that we address ourselves in this report.

Our experimental material consisted of 334 papers in Vol. 112 (1958) of the *Physical Review*. These papers were arranged into groups according to two schemes: the Analytic Subject Index (ASI) and Bibliographic Coupling (BC). In the first instance, the editors supply a list of 73 subject categories. Someone in the editorial office considers each paper separately and places it into one or more categories. At the end of this process, each of the 334 papers finds itself in a subject category. Groups of papers are thus formed under the heading of each subject category. There will be as many groups as there are categories, in this case 73, although some of the groups may be empty. In the second case (BC) each paper is compared to each of the remaining 333. Those that share one or more bibliographic items with the test paper are considered to be members of a group. There will be as many groups as there are papers in the sample, in this case 334, although some of the groups may be empty.

Table 1 shows the group-forming characteristics of the two methods.

Although there is no upper limit to the number of categories in which a given paper may be placed, we find that the practitioners of ASI in the *Physical Review* try to keep this number well below four. Of the 334 papers, 159 are in one category only, 124 find themselves in two groups, 47 in three groups, and only four papers are members of four groups. In the case of BC, the distribution of papers among groups is the same as the distribution of groups having N papers (third column in Table 1); 44 papers fall into no group at all, 56 in one group only, 45 in two groups, 30 in three groups, etc. We see that in ASI a collection of papers will be distributed among a smaller number of groups and each paper will be a member of fewer groups than in BC. This is of no particular significance at this time since both methods generate enough groups to allow comparison. We must, however, point out the following significant differences in the mechanics of group formation because they will have to be appealed to later when we examine the results of the two methods.

In the case of BC, the process for group formation follows a fixed prescription. Even when performed by a human being, it is completely mechanical. There is no bending or twisting to make a fit; no discretion or leeway

Table 1. Groups formed according to ASI and BC.

Number of papers in the group (N)	Number of groups having N papers ASI	BC
0	10	44
1	9	56
2	5	45
3	6	30
4	7	31
5	5	17
6	3	18
7	2	11
8	5	6
9	5	14
10	5	13
11	0	10
12	0	10
13	2	6
14	0	4
15	2	6
16	2	5
17	1	2
18	0	1
20	1	1
21	1	1
23	1	0
24	0	1
25	0	1
26	0	1
27	1	0
33	1	0
39	1	0
49	1	0

whatsoever is available to the operator. The coupling criterion is fixed and the results were determined once and for all by the authors who wrote the paper. Not so in the case of ASI. Here much is left to the discretion of the editors and indexers. The very choice of categories reflects the expectation of the editors. In areas where large numbers of papers are expected, a much finer subdivision of field categories is provided than in areas where the expectation of papers is low. Thus we have a class "Liquid Helium" referring to a particular state of one element and a class "Nuclear Reactions Induced by Neutrons" equally narrow in scope, while at the same time we have "Biophysics," "Geophysics," and "Acoustics," which cover major scientific fields. This reflects the editors' expectation that there will be very few papers in Biophysics, Geophysics, and Acoustics in the *Physical Review*. Or consider the class "Nuclear Reactions—General" in addition to five classes for nuclear reactions each induced by a specific particle. Clearly the editors take a pragmatic view of the field and divide it into 73 cells of equal expectation rather than 73 logical subdivisions. This leads to a reasonable distribution of papers among cells with a minimum of empty categories. It also leads to a lack of uniformity in the relatedness of two papers that find themselves in one cell. For example, two papers assigned to the category (or cell) Liquid Helium would be related much more closely to each other than two papers assigned to the category Astrophysics.

In the case of BC, all groups should exhibit the same degree of relatedness. But since this relatedness is statistical, there is in actual practice a distribution of relatedness *within* each group, although the properties from group to group are more or less uniform.

Before proceeding with the actual comparison of the groups formed by the two methods, we remind the reader that in this test we are concerned merely with the mechanics of group formation and not with its meaning or relevance to retrieval. If we find an equivalence between the two sets of groups, they may be equally good, equally bad or equally indifferent as far as a given mode of retrieval is concerned. In other words, we address ourselves to questions of the first order. We do this in order to delineate clearly the problem of this report and to avoid confusion. We fully recognize that questions of the second order are of prime significance. But they must be postponed. Since, however, all the papers in our illustrations are in Vol. 112 of the *Physical Review* and are identified by page number, the reader is free to refer to them if he so desires and form his own judgment as to their relevance to some retrieval application that he may have in mind.

The comparison will be performed in four stages:

1. Given the papers in a group according to the BC criterion, how are they regarded by the ASI criterion?
2. Given the papers in a group according to the ASI criterion, how are they regarded by the BC criterion?
3. Given two papers considered to be strongly related by BC, what is the verdict of ASI?

4. Given two papers considered to be strongly related by ASI, what is the verdict of BC?

• First Stage

Each of the 334 papers in Vol. 112 was used as a test paper to generate a group according to BC. For example, the paper on page 624 acting as P_0 generated a group of fourteen papers as shown in the first row in Table 2 under G_A . We then consulted Vol. 112 of the *Physical Review* to see how each of these papers was classified according to ASI. This is indicated by an X in the appropriate row. Thus we see that paper 273 was placed in three ASI categories, paper 614 in two, etc. For the group as a whole, we find that, according to the ASI criterion, all but one belonged together in the category headed "Elementary Particle Interactions." Ten of the fourteen papers were judged to belong in the category "Mesons and Hyperons." Although seven of the available 73 ASI categories were used to describe the fourteen papers in this G_A , two of these categories would have accounted for the entire list with some redundancy to spare. This G_A shows a particularly high homogeneity with respect to the judgment of ASI. At the other extreme, however, consider the group generated by $P_0=968$ as shown in Table 3. There are nine papers in this group and, in order to account for all of them, we need six of the eight ASI categories used. No single ASI category contains more than two of the nine papers and there is very little redundancy. A formal appraisal of Table 3 would suggest a low correlation between the results of BC and ASI. However, an examination of the ASI categories shows that this only slightly correlated case is not

Table 2. First stage comparison of BC and ASI ($P_0=624$; see text for explanation).

ASI Categories	G_A													
	273	605	614	642	979	981	986	1303	1311	1793	1813	2053	2107	2135
Elementary particle interactions	X	—	X	X	X	X	X	X	X	X	X	X	X	X
Mesons and hyperons	X	X	X	—	X	X	—	—	X	X	—	X	X	X
Field theory	X	—	—	X	—	—	—	—	—	—	X	—	—	X
Nuclear photoeffects	—	—	—	—	—	—	X	—	—	—	—	X	—	—
Scattering of mesons and hyperons	—	—	—	—	—	—	X	—	—	—	X	—	—	—
Scattering of protons	—	—	—	—	—	—	—	X	—	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	—	—	—	—	—	X	—	—	—	—	—

Table 3. First stage comparison of BC and ASI ($P_0=968$; see text for explanation).

ASI Categories	G_A							
	232	590	906	1285	1325	1658	1667	1719
Mesons and hyperons	—	X	—	—	—	—	—	X
Scattering of mesons and hyperons	—	—	—	—	X	—	—	—
Nuclear reactions induced by deuterons and tritons	—	—	—	—	—	—	X	—
Scattering of protons	—	—	—	—	—	X	—	—
Nuclear spectra	X	—	—	—	—	—	—	—
Nuclear structure theory	—	—	X	—	—	—	—	X
Mathematical methods	—	—	X	—	—	—	—	—
Nuclear reactions induced by protons	—	—	—	X	—	X	—	—

as bad as it looks because the ASI categories are logically very close to each other, and a slight change in emphasis by the indexer could have changed the picture materially. Similar tables were made for each of the 334 groups generated by the method of BC. To keep this report to manageable size, we shall not reproduce all the data. We give a sizeable sample of the results and a statistical summary of the whole. Table 4 shows the comparison of five groups, each consisting of ten papers. Table 5 shows the results of five groups each containing fifteen papers. Table 6 shows five groups of 20 or more papers. Table 7 is a statistical summary of the results of all of the 334 papers.

To obtain an empirical measure of how the results of the two methods compare, we consider that if all the members of G_A could be accounted for by three ASI categories there is a good match between the two methods. When that is not so we use the additional measure of the fraction of ASI categories necessary to account for all the members of the group.

The summary data (Table 7) are given in the following way. Each paper has four numbers associated with it. These numbers are given as two fractions A/B and C/D where:

A—the largest number of papers in G_A included in three ASI categories.

B—the number of papers in G_A .

C—the number of ASI categories that would account for all the papers in G_A .

D—the total number of ASI categories used to describe all the papers in G_A .

A little consideration will show that a "good" correlation between the results of BC and ASI will be characterized by a relatively large value of A/B and a small value of C/D . The ratio of the two fractions may be used as a measure of the correlation. Thus in the example of Table 2 this ratio is $\frac{14}{7} / \frac{2}{7} = 3.5$; in the example of Table 3 the ratio is $\frac{4}{4} / \frac{2}{8} = 9$. We do not attach any theoretical significance to this ratio. It is used here as a matter of convenience.

In evaluating the results of this comparison as summarized in the table above, we eliminated papers that formed BC groups of three or less. This was necessary because we happened to pick three ASI categories as a reasonable measure of agreement between the results of BC and ASI. The average value of A/B for all BC groups having four papers or more is 0.9. This means that, on the average, the a priori verdict of ASI indexing is that three ASI categories will account for nine out of every ten papers in the BC groups. Had we not eliminated the smaller groups this number would have been even larger. For the same group of papers the average value of the fraction C/D is 0.5. This means that one-half of all the ASI categories used by the editors to

index the papers in each BC group would have sufficed to account for all the members of the group.

• Second Stage

In this experiment, we start with the papers found under a given ASI heading and call them a group. We then observe the distribution of the members of this group among the BC groups. It will be recalled that the nature of BC is such that every paper generates a group. If, therefore, we find that a certain ASI heading contains 15 papers, we expect each of these to generate a BC group. The papers in the ASI group will thus be distributed among 15 BC groups. The question that we must answer is—Are these groups more alike than one might expect from random association? A measure of this answer might be to note how many of the 15 BC groups are necessary to account for all of the 15 papers in the single ASI group. An example will help clarify what we mean. Consider the group of papers under the ASI heading "Elementary Particle Interactions." There are 36 papers in this group. These are shown in the first column of Table 8. Each of these papers generates a group by the BC method. Thus paper 624 generates a group that includes, among others, 13 of the 36 papers in column 1. These are marked X in column 2. Paper 986 generates a BC group that includes 12 of the 36 papers. These are marked X in column 3. These two BC groups account for 19 of the 36 papers in the given ASI category. If we continue this process we find that 8 BC groups account for 33 of the 36 papers. The remaining papers (587, 1335 and 1642) do not form groups according to BC. This is a highly correlated example. A less correlated example is given by the ASI category "Crystalline State," Table 9. There are 45 papers in this category. The BC groups generated by seven of these account for 21. Beyond that no single BC group contributes more than one paper to the category. The data in general fall between these two extremes, depending on the "logical size" of the ASI categories.

As was pointed out previously in this report, the ASI categories were not chosen entirely on the basis of a logical subdivision of the subject field. Allowances were made for the expected number of papers in each category. The amount of relatedness that one may expect within each ASI category varies. It is therefore to be expected that when we apply a fixed measure of relatedness to the papers in each category, the results will show extreme variations. Some of the tight categories like "Gamma Rays" or "Properties of Films" will gather papers highly related and these will show great correlation when compared to the results of BC. On the other hand, categories like "Geophysics," "Mechanics" and "Biophysics," to say nothing of "Miscellaneous," will gather a few papers clearly to be included in the subject categories, but not necessarily related.

Table 4 First stage comparison of BC and ASI: five groups of ten papers each.

(a) $P_o = 344$	
ASI Categories	G_A
	31 90 186 322 334 812 935 1092 1137 1169
Crystalline state	— X — X X X — X X —
Superconductivity	X — — — — — — — —
Dielectrics and dielectrical properties	— X — — — — — — —
Atomic and molecular beams	— — X — — — — — —
Magnetic resonance	— — X — — — X — — X
Nuclear moments and spin	— — X — — — X — — —
Elasticity and plasticity	— — — — — — X — —
Spectra, general	— — — — — — — — X
(b) $P_o = 1227$	
ASI Categories	G_A
	503 547 897 931 1224 1658 1975 1989 2020 2029
Nuclear reactions, general	— — — — — — — — X
Nuclear reactions induced by deuterons and tritons	X — — — X — X — X —
Nuclear reactions induced by α particles and He^3	— X — X — — — — —
Nuclear reactions induced by protons	— — X — — X — — —
Nuclear spectra	— — — X — — — — —
Scattering of protons	— — — — — X — — —
Scattering of α particles	— — — — — — X — —
Nuclear structure theory	— — — — — — — — X
Scattering, general	— — — — — — — — X
(c) $P_o = 1311$	
ASI Categories	G_A
	271 614 624 903 1027 1287 1295 1303 1313 1802
Elementary particles	— X X — X X — X — X
Mesons and hyperons	— X X — — — — — —
Nuclear reactions induced by protons	— — — — — X — X —
Scattering, general	X — — — — — — — —
Mathematical methods	— — — X — — — — —
Nuclear structure theory	— — — X — — — — —
Field theory	— — — — X — — — X
Scattering of mesons and hyperons	— — — — X — — — X
Scattering of protons	— — — — — X — — —
(d) $P_o = 1900$	
ASI Categories	G_A
	31 328 812 903 994 1008 1056 1083 1522 1888
Superconductivity	X — — — — — — X X X
Crystalline state	— X X — — — — — —
Gases	— X — — — — — — —
Magnetic properties	— X — — — — — — —
Mathematical methods	— — — X — — — — —
Nuclear structure theory	— — — X X X — — —
Field theory	— — — — X X — — —
Statistical mechanics	— — — — X X — — —
Thermal properties	— — — — — X — — —
(e) $P_o = 2020$	
ASI Categories	G_A
	503 890 903 923 1227 1658 1667 1975 2029 2048
Nuclear reactions induced by deuterons and tritons	X — — X — — X X — —
Nuclear photoeffects	— X — — — — — — —
Nuclear spectra	— — X — — — — — —
Nuclear structure theory	— — — X — — — — X
Nuclear reactions, general	— — — — X — — — X
Nuclear reactions induced by protons	— — — — X — — — —
Scattering of protons	— — — — — — — — —
Scattering, general	— — — — — — — — X
Scattering of α particles	— — — — — — — — X

Tabl. 5. First stage comparison of BC and ASI. five groups of fifteen papers each.

(a) P _a = 503		G _A														
ASI Categories		232	452	461	903	923	931	1200	1227	1252	1337	1375	1703	1975	2020	2029
Nuclear reactions induced by deuterons and tritons		—	—	X	—	X	—	—	—	—	—	—	—	X	X	—
Nuclear spectra		X	—	—	X	—	X	—	—	X	—	—	—	—	—	—
Nuclear structure theory		—	X	—	—	X	—	—	—	—	—	—	—	—	—	X
Nuclear reactions induced by α particles and He ³		—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	X	—	X	—	—	—	—	—	—
Nuclear reactions, general		—	—	—	—	—	—	—	X	—	—	—	—	—	—	X
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	X	—	X	—	—	—
Elementary particles		—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Field theory		—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
(b) P _a = 554		G _A														
ASI Categories		461	468	518	560	568	897	954	963	1183	1267	1295	1679	1730	1994	2026
Nuclear photoeffects		—	—	—	X	—	—	—	—	—	—	—	—	—	X	—
Nuclear reactions induced by deuterons and tritons		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra		X	X	—	—	—	—	—	—	—	—	—	—	X	—	X
Scattering of electrons and positrons		—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	X	—	—	—	—	X	X	—	—	—
Nuclear structure theory		—	—	—	—	—	—	X	—	—	—	—	—	—	—	—
Elementary particles		—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Neutrinos and antineutrinos		—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Atomic structure and spectra		—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
X-rays		—	—	—	—	—	—	—	X	X	—	—	—	—	—	—
Radiation		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
(c) P _a = 897		G _A														
ASI Categories		513	532	554	560	935	1183	1227	1247	1658	1672	1702	1730	1958	2020	2029
Nuclear reactions induced by protons		—	X	—	—	—	—	—	—	X	Y	X	—	—	—	—
Nuclear spectra		X	—	—	—	—	—	—	X	—	—	—	X	X	X	—
X-rays		—	X	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear photoeffects		—	—	X	X	—	—	—	—	—	—	—	—	—	—	—
Magnetic resonance		—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Nuclear moment and spin		—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Atomic structure and spectra		—	—	—	—	—	X	—	—	—	—	—	—	—	—	—
Nuclear reactions, general		—	—	—	—	—	—	X	—	—	—	—	—	—	—	X
Scattering of protons		—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
(d) P _a = 1027		G _A														
ASI Categories		614	669	866	966	994	1008	1287	1311	1344	1370	1375	1380	1555	1613	1543
Elementary particles		X	—	—	X	—	—	X	X	X	—	X	X	—	X	—
Field theory		—	X	—	—	X	X	—	X	—	—	X	X	—	X	X
Mesons and hyperons		X	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Magnetic resonance		—	—	X	—	—	—	—	—	—	—	—	—	—	—	—
Molecular structure and spectra		—	—	X	—	—	—	—	—	—	—	—	—	—	—	—
Scattering of mesons and hyperons		—	—	—	X	—	—	X	—	—	X	—	—	—	X	—
Nuclear structure theory		—	—	—	—	X	X	—	—	—	—	—	—	—	—	—
Statistics' mechanics		—	—	—	—	X	X	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Crystalline state		—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Dielectrics and dielectrical properties		—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Quantum mechanics		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
Radiation		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

Table 5—Continued

(e) $P_o = 2135$		G_A													
ASI Categories		267	273	580	605	624	642	665	979	981	986	1375	1746	1793	2053
Elementary particles		X	X	—	—	X	X	X	X	X	X	X	—	X	X
Field theory		—	X	—	—	—	X	—	—	—	—	X	—	—	—
Mesons and hyperons		X	X	X	X	X	—	X	X	X	—	—	X	X	X
Quantum electrodynamics		X	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear photoeffects		—	—	—	—	—	—	—	—	X	—	—	—	—	X
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	X	—	—	—	—

Table 6. First stage comparison of BC and ASI: five groups of twenty or more papers each.

(a) $P_o = 232$		G_A																
ASI Categories		244	267	273	452	461	503	580	605	665	877	893	968	979	1240	1252	1337	1375
Elementary particles		X	X	X	—	—	—	—	—	X	—	—	—	X	—	—	—	X
Neutrinos and anti-neutrinos		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra		X	—	—	—	—	—	—	—	—	X	—	—	—	X	X	—	—
Mesons and hyperons		—	X	X	—	—	—	X	X	X	—	X	X	X	—	—	—	—
Quantum electrodynamics		—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Field theory		—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	X	—
Nuclear structure theory		—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by deuterons and tritons		—	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—	—
Radioactivity		—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Scattering of protons		—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—
Nuclear reactions induced by mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	X	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Nuclear reactions, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X

(b) $P_o = 273$		G_A													
ASI Categories		232	244	481	580	605	614	624	665	877	893	945	963	979	1240
Nuclear spectra		X	X	X	—	—	X	X	X	X	X	X	—	—	X
Elementary particles		—	X	—	—	—	—	—	—	—	—	—	—	—	X
Neutrinos and anti-neutrinos		—	X	—	—	—	—	—	—	—	—	—	—	—	X
Mesons and hyperons		—	—	—	X	X	X	X	X	—	—	—	—	—	X
Radioactivity		—	—	—	—	—	—	—	—	X	—	—	—	—	—
Scattering of protons		—	—	—	—	—	—	—	—	—	X	—	—	—	—
Field theory		—	—	—	—	—	—	—	—	—	—	—	—	—	X
Atomic structure and spectra		—	—	—	—	—	—	—	—	—	—	—	—	—	X
Electrons and positrons		—	—	—	—	—	—	—	—	—	—	—	—	—	X
Quantum electrodynamics		—	—	—	—	—	—	—	—	—	—	—	—	—	X

(c) $P_o = 605$		G_A													
ASI Categories		232	267	273	580	614	624	665	979	986	1021	1282	1335	1375	1658
Mesons and hyperons		—	X	X	X	X	X	X	X	—	—	—	—	X	—
Nuclear spectra		X	—	—	—	—	—	—	—	—	—	—	—	—	—
Elementary particles		—	X	X	—	X	X	X	X	X	—	—	X	—	—
Quantum electrodynamics		—	X	—	—	—	—	—	—	—	—	—	—	—	—
Field theory		—	—	X	—	—	—	—	—	—	X	—	—	—	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	X	—	—	—	X	—
Scattering of electrons and positrons		—	—	—	—	—	—	—	—	—	—	X	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	—	—	—	—	—	X	—
Scattering of protons		—	—	—	—	—	—	—	—	—	—	—	—	X	—
Nuclear structure theory		—	—	—	—	—	—	—	—	—	—	—	—	X	—

Table 6—Continued

(d) P ₀ = 1375		C _A																
ASI Categories		232	267	273	452	461	503	580	605	665	931	945	979	1027	1252	1337	1637	1746
		1763	1796	1813	1975	2010	2029	2128	2135									
Elementary particles		—	X	X	—	—	—	—	—	X	—	—	X	X	—	—	X	X
Field theory		—	—	X	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Nuclear spectra		X	—	—	—	—	—	—	—	—	X	X	—	—	X	—	—	—
Mesons and hyperons		—	X	X	—	—	—	X	X	X	—	—	X	—	—	—	X	—
Quantum electrodynamics		—	X	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—
Nuclear structure theory		—	—	—	X	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by deuterons and tritons		—	—	—	—	X	X	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by α particles and He ³		—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—
Atomic structure and spectra		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—
Electrons and positrons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—
Nuclear reactions, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

(e) P ₀ = 2020		G _A																
ASI Categories		186	232	452	461	503	518	855	897	903	923	931	986	1192	1227	1252	1337	1376
		1658	1672	1763	1958	1975	1981	1989	2020	2048								
Atomic and molecular beams		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Magnetic resonance		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear moment and spin		X	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear spectra		—	X	—	—	—	X	—	—	X	—	—	—	—	—	X	—	—
Nuclear structure theory		—	—	X	—	—	—	—	—	—	X	—	—	—	—	—	—	—
Nuclear reactions induced by deuterons and tritons		—	—	—	X	X	—	—	—	—	X	X	—	—	—	—	—	—
Atomic structure and spectra		—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Spectra, general		—	—	—	—	—	—	X	—	—	—	—	—	—	—	—	—	—
Nuclear reactions induced by protons		—	—	—	—	—	—	—	X	—	—	—	—	—	X	—	—	—
Elementary particles		—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—
Scattering of mesons and hyperons		—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—
Nuclear reactions induced by neutrons		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Scattering of neutrons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Nuclear reactions, general		—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—
Field theory		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Scattering of protons		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Scattering, general		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X
Scattering of α particles		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Table 7. Statistical summary of first stage comparison.
Page numbers refer to title page of articles in Vol. 112.
See text for explanation of the variables A/B and C/D.
See also remarks following this table.

Page	A/B	C/D	Page	A/B	C/D
31	7/7	3/6	369	4/4	1/4
49	4/4	3/5	381	4/4	1/6
80	4/4	3/4	397	4/4	1/4
136	5/5	3/5	452	9/13	5/11
179	5/5	3/4	461	11/17	3/13
182	4/4	3/4	468	5/10	5/6
186	12/13	3/6	503	10/15	7/10
191	5/6	4/6	512	4/4	3/4
203	4/5	4/5	532	10/12	5/7
229	4/4	1/2	547	6/7	4/6
262	5/6	3/6	554	3/15	5/11
267	11/12	3/6	560	5/11	6/6
271	5/5	3/6	580	13/16	5/11
298	5/6	3/5	605	17/21	7/10
317	5/6	3/6	614	3/6	3/5
322	5/5	1/2	618	9/6	2/7
326	4/4	4/4	624	14/14	3/7
328	4/4	4/4	642	13/12	3/6
334	4/4	1/1	665	12/12	3/7
337	4/4	3/4	635	3/8	3/7
344	19/10	3/6	638	5/4	2/5

Table 7—Continued

Page	A/B	C/D	Page	A/B	C/D
715	$\frac{5}{6}$	$\frac{3}{6}$	1504	$\frac{4}{4}$	$\frac{3}{6}$
732	$\frac{9}{6}$	$\frac{3}{6}$	1522	$\frac{5}{4}$	$\frac{5}{6}$
812	$\frac{1}{1} \frac{1}{6}$	$\frac{1}{1} \frac{1}{6}$	1599	$\frac{4}{4}$	$\frac{1}{4}$
828	$\frac{4}{4}$	$\frac{3}{6}$	1603	$\frac{5}{6}$	$\frac{1}{6}$
843	$\frac{5}{6}$	$\frac{4}{4}$	1620	$\frac{4}{4}$	$\frac{1}{4}$
855	$\frac{5}{4}$	$\frac{3}{6}$	1637	$\frac{5}{6}$	$\frac{2}{4}$
870	$\frac{4}{4}$	$\frac{3}{6}$	1658	$\frac{9}{4} \frac{6}{6}$	$\frac{9}{1} \frac{3}{3}$
890	$\frac{4}{4}$	$\frac{2}{3}$	1667	$\frac{5}{6}$	$\frac{4}{4}$
897	$\frac{1}{1} \frac{1}{1} \frac{5}{5}$	$\frac{5}{1} \frac{0}{0}$	1672	$\frac{1}{6}$	$\frac{5}{6}$
906	$\frac{1}{1} \frac{1}{1} \frac{5}{5}$	$\frac{1}{1} \frac{2}{2}$	1719	$\frac{5}{1} \frac{0}{0}$	$\frac{5}{1} \frac{3}{3}$
923	$\frac{5}{6}$	$\frac{4}{6}$	1730	$\frac{5}{6}$	$\frac{4}{6}$
931	$\frac{5}{4}$	$\frac{5}{6}$	1746	$\frac{1}{1} \frac{2}{2}$	$\frac{3}{5}$
945	$\frac{1}{1} \frac{2}{2}$	$\frac{2}{4}$	1759	$\frac{4}{4}$	$\frac{3}{4}$
954	$\frac{5}{4}$	$\frac{5}{6}$	1763	$\frac{9}{1} \frac{4}{4}$	$\frac{1}{1} \frac{2}{2}$
963	$\frac{7}{8}$	$\frac{4}{4}$	1773	$\frac{9}{1} \frac{1}{1}$	$\frac{9}{8}$
968	$\frac{6}{1} \frac{0}{0}$	$\frac{1}{1} \frac{0}{0}$	1793	$\frac{5}{6}$	$\frac{5}{6}$
979	$\frac{1}{1} \frac{2}{2}$	$\frac{3}{5}$	1802	$\frac{5}{5}$	$\frac{3}{4}$
981	$\frac{7}{4}$	$\frac{1}{4}$	1813	$\frac{1}{1} \frac{3}{3}$	$\frac{3}{6}$
986	$\frac{1}{1} \frac{4}{4} \frac{6}{6}$	$\frac{2}{1} \frac{1}{1}$	1877	$\frac{5}{6}$	$\frac{3}{5}$
992	$\frac{5}{4}$	$\frac{1}{4}$	1888	$\frac{6}{6}$	$\frac{3}{4}$
994	$\frac{9}{1} \frac{5}{5}$	$\frac{9}{1} \frac{5}{5}$	1900	$\frac{9}{1} \frac{0}{0}$	$\frac{4}{1} \frac{0}{0}$
1008	$\frac{6}{1} \frac{0}{0}$	$\frac{1}{1} \frac{3}{3}$	1950	$\frac{5}{5}$	$\frac{3}{5}$
1021	$\frac{1}{1} \frac{0}{0} \frac{3}{3}$	$\frac{9}{9}$	1954	$\frac{1}{1} \frac{0}{0} \frac{0}{0}$	$\frac{3}{6}$
1027	$\frac{1}{1} \frac{3}{3} \frac{5}{5}$	$\frac{5}{1} \frac{4}{4}$	232	$\frac{1}{1} \frac{2}{2} \frac{5}{5}$	$\frac{1}{1} \frac{5}{5}$
1056	$\frac{5}{6}$	$\frac{4}{8}$	273	$\frac{2}{0} \frac{2}{0}$	$\frac{3}{1} \frac{0}{0}$
1083	$\frac{5}{6}$	$\frac{3}{5}$	1975	$\frac{5}{1} \frac{7}{7}$	$\frac{1}{0} \frac{1}{1} \frac{3}{3}$
1092	$\frac{9}{6}$	$\frac{3}{5}$	1981	$\frac{4}{4}$	$\frac{3}{4}$
1112	$\frac{4}{4}$	$\frac{3}{6}$	1989	$\frac{6}{4}$	$\frac{4}{4}$
1130	$\frac{4}{4}$	$\frac{1}{4}$	1994	$\frac{5}{5}$	$\frac{3}{4}$
1151	$\frac{4}{4}$	$\frac{2}{4}$	2010	$\frac{4}{4}$	$\frac{3}{4}$
1157	$\frac{5}{6}$	$\frac{3}{6}$	2020	$\frac{1}{1} \frac{0}{0}$	$\frac{6}{6}$
1183	$\frac{7}{9}$	$\frac{5}{8}$	2043	$\frac{5}{6}$	$\frac{4}{6}$
1192	$\frac{9}{1} \frac{2}{2}$	$\frac{9}{1} \frac{0}{0}$	2048	$\frac{5}{1}$	$\frac{1}{1} \frac{0}{0}$
1200	$\frac{9}{1} \frac{0}{0}$	$\frac{4}{4}$	2053	$\frac{9}{9}$	$\frac{3}{6}$
1210	$\frac{4}{5}$	$\frac{4}{6}$	2072	$\frac{5}{5}$	$\frac{3}{6}$
1217	$\frac{5}{4}$	$\frac{5}{6}$	2077	$\frac{1}{1} \frac{0}{0} \frac{4}{4}$	$\frac{1}{1} \frac{1}{1}$
1227	$\frac{9}{1} \frac{0}{0}$	$\frac{3}{6}$	2112	$\frac{4}{4}$	$\frac{3}{5}$
1247	$\frac{7}{4}$	$\frac{3}{6}$	2128	$\frac{9}{6}$	$\frac{3}{5}$
1252	$\frac{8}{1} \frac{6}{6}$	$\frac{8}{1} \frac{4}{4}$	2135	$\frac{6}{6}$	$\frac{1}{4}$
1257	$\frac{3}{8}$	$\frac{3}{5}$	2026	$\frac{9}{1} \frac{1}{1}$	$\frac{3}{6}$
1267	$\frac{5}{5}$	$\frac{2}{6}$	2029	$\frac{1}{1} \frac{4}{4} \frac{6}{6}$	$\frac{1}{0} \frac{1}{1} \frac{8}{8}$
1282	$\frac{6}{6}$	$\frac{6}{1} \frac{0}{0}$	1169	$\frac{6}{6}$	$\frac{2}{5}$
1287	$\frac{8}{8}$	$\frac{3}{4}$	1627	$\frac{5}{1} \frac{0}{0}$	$\frac{4}{1} \frac{3}{3}$
1295	$\frac{9}{1} \frac{1}{1}$	$\frac{1}{1} \frac{2}{2}$	1958	$\frac{5}{1} \frac{2}{2}$	$\frac{6}{1} \frac{2}{2}$
1303	$\frac{9}{6}$	$\frac{3}{1} \frac{0}{0}$	518	$\frac{8}{1} \frac{4}{4}$	$\frac{9}{1} \frac{2}{2}$
1311	$\frac{9}{1} \frac{0}{0}$	$\frac{4}{9}$	481	$\frac{1}{1} \frac{2}{2} \frac{2}{2}$	$\frac{3}{1} \frac{0}{0}$
1319	$\frac{7}{4}$	$\frac{3}{6}$	489	$\frac{4}{4}$	$\frac{3}{5}$
1337	$\frac{8}{1} \frac{1}{1}$	$\frac{5}{6}$	903	$\frac{1}{1} \frac{0}{0}$	$\frac{5}{1} \frac{0}{0}$
Page	A/B	C/D	935	$\frac{8}{9}$	$\frac{4}{6}$
1344	$\frac{9}{1} \frac{0}{0}$	$\frac{4}{4}$	877	$\frac{9}{6}$	$\frac{3}{6}$
1367	$\frac{4}{4}$	$\frac{2}{4}$	893	$\frac{1}{1} \frac{1}{1} \frac{1}{1}$	$\frac{2}{6}$
1370	$\frac{7}{4}$	$\frac{3}{5}$	1240	$\frac{1}{1} \frac{1}{1} \frac{1}{1}$	$\frac{3}{6}$
1375	$\frac{1}{1} \frac{3}{3} \frac{5}{5}$	$\frac{9}{1} \frac{4}{4}$	1684	$\frac{1}{1} \frac{0}{0}$	$\frac{3}{6}$
1391	$\frac{9}{4}$	$\frac{4}{1} \frac{3}{3}$	1702	$\frac{8}{1} \frac{1}{1}$	$\frac{5}{8}$
1404	$\frac{5}{5}$	$\frac{2}{5}$	1998	$\frac{1}{1} \frac{1}{1} \frac{1}{1}$	$\frac{2}{6}$
1429	$\frac{6}{6}$	$\frac{3}{5}$	2004	$\frac{4}{8}$	$\frac{3}{5}$
1456	$\frac{4}{5}$	$\frac{3}{6}$	2061	$\frac{1}{1} \frac{3}{3}$	$\frac{4}{8}$
1488	$\frac{4}{4}$	$\frac{3}{4}$	481	$\frac{1}{1} \frac{2}{2}$	$\frac{3}{1} \frac{0}{0}$

In order to analyze the results of this comparison statistically, we must have a measure of relatedness for the ASI categories. This would lead us into a measure of ASI independent of its comparison to BC, a task that we do not undertake at this time. Of course, one could claim that the extent of correlation with the results of BC is itself a measure of the "size" of the ASI category. Such an attitude, however, cannot properly be maintained in the course of an experiment that involves BC as a variable. We, therefore, draw the qualitative

conclusion that in those cases where the ASI categories seem to be of a small enough "logical size," the correlation with BC is good. When the "logical size" of the categories broadens to a point where sizeable fields of knowledge are involved, the correlation diminishes.

• Third Stage

At this time it will profit the reader to review the defining statement on Bibliographic Coupling as given in

Table 8. Second stage comparison of BC and ASI.
Column A=papers in the category "Elementary Particle Interactions."

Subsequent columns are the BC groups generated by the papers indicated at the top of the column.
See text for explanation and comments.

A	624	986	1813	2135	273	642	1344	981
244	--	--	--	--	X	--	--	--
267	--	--	--	X	--	--	--	--
373	X	--	--	X	--	--	--	--
283	--	--	--	--	--	X	--	--
587	--	--	--	--	--	--	--	--
614	X	--	--	--	X	--	--	--
618	--	--	X	--	--	--	--	--
622	--	--	--	--	--	--	--	X
624	--	X	X	X	X	X	--	X
642	X	X	X	X	--	--	X	X
665	--	--	--	X	X	--	--	--
963	--	--	--	X	X	--	--	--
979	X	--	--	X	X	--	--	--
981	X	X	--	X	--	X	--	--
986	X	--	X	X	--	X	X	X
992	--	--	--	--	--	--	X	--
1021	--	X	X	--	--	X	--	--
1027	--	X	X	--	--	--	X	--
1287	--	X	X	--	--	--	X	--
1303	X	--	X	--	--	--	--	--
1311	X	--	--	--	--	--	--	--
1335	--	--	--	--	--	--	--	--
1344	--	X	X	--	--	X	--	--
1375	--	--	X	X	X	--	--	--
1380	--	--	--	--	--	--	X	--
1637	--	--	--	--	X	--	--	--
1642	--	--	--	--	--	--	--	--
1793	X	X	--	X	--	X	--	X
1802	--	--	X	--	--	--	--	--
1813	X	X	--	--	--	X	X	--
2063	X	X	--	X	--	X	--	X
2061	--	--	--	--	X	--	--	--
2077	--	X	X	--	--	--	--	--
2107	X	--	--	--	--	--	--	--
2123	--	--	--	X	--	--	--	--
2135	X	X	--	--	X	X	--	X

the footnote on the first page. The statement indicates two conditions of relatedness between papers. A group of papers may be related by virtue of each being a member of a given G_A , but also each member of G_A forms a pair relation with P_0 . The first instance produces groups of papers; the second gives pairs of papers. The first and second stages of comparison considered how the various G_A groups compare with groups of papers formed by assignment to ASI categories. In the present comparison, we consider the pairs of papers formed between P_0 and each member of G_A .

The number of pairs formed is equal to one-half of the cumulative total of the papers in all the G_A groups. In Vol. 112 of the *Physical Review*, 876 pairs of papers were formed variously coupled according to the indicated coupling strength shown in Table 10.

Since there are very few pairs of high coupling strength we place all those with $n = 5$ and larger into one group.

* The random pairs were chosen by picking 386 telephone numbers, two at a time, from a random place in the directory. The nearest page numbers corresponding to the telephone numbers were called a random pair of papers.

Table 9. Second stage comparison of BC and ASI.
Column A=papers in the category "Crystalline State."

Subsequent columns are the BC groups generated by the papers indicated at the top of the column.
See text for explanation and comments.

A	344	715	1092	1603	322	381	1620
54	--	--	--	--	--	--	--
64	--	--	--	--	--	--	--
73	--	--	--	--	X	--	--
80	--	X	--	--	--	X	--
90	X	--	--	--	--	--	--
103	--	--	--	--	--	--	--
322	X	--	X	--	--	--	--
328	--	--	--	--	--	--	--
334	X	--	--	--	X	--	--
337	--	--	--	--	--	--	X
344	--	--	X	--	X	--	--
381	--	X	--	--	X	--	--
685	--	--	--	--	--	--	--
698	--	X	--	--	--	--	--
708	--	X	--	--	--	--	--
715	--	--	--	--	--	X	--
722	--	--	--	--	--	--	X
725	--	--	--	--	--	--	--
751	--	--	--	--	--	--	--
765	--	--	--	--	--	--	--
785	--	--	--	--	--	--	--
804	--	--	--	--	--	--	--
812	X	--	--	--	--	--	--
837	--	--	--	--	--	--	--
1058	--	--	--	--	--	--	--
1061	--	--	--	--	--	--	--
1063	--	--	--	--	--	--	--
1092	X	--	--	--	X	--	--
1117	--	--	--	--	--	--	--
1130	--	--	--	X	--	--	--
1157	X	--	--	--	--	--	--
1516	--	--	--	X	--	X	--
1534	--	--	--	--	--	--	--
1540	--	--	--	--	--	--	--
1555	--	--	--	--	--	--	--
1567	--	--	X	--	--	--	--
1571	--	--	--	--	--	--	--
1587	--	--	--	--	--	--	--
1603	--	--	--	--	--	--	--
1616	--	--	--	--	--	--	--
1620	--	--	--	--	--	--	--
1857	--	--	--	--	--	--	--
1863	--	--	X	--	--	--	--
1877	--	--	--	--	--	X	--
1917	--	--	--	X	--	--	--

Table 10. Pairs of papers and their coupling strength

635 pairs coupled with	1 coupling unit
132 "	2 "
47 "	3 "
25 "	4 "
13 "	5 "
9 "	6 "
4 "	7 "
5 "	8 "
1 "	9 "
3 "	10 "
1 "	11 "
1 "	12 "

Those with $n = 3$ and 4 form another group of pairs. The third and fourth groups consist of pairs with $n = 2$ and $n = 1$ respectively. As a control we chose a random group of 100 pairs.* We then examined how each of the

976 pairs was regarded by the ASI method of indexing. We considered that if the pair shares one ASI category a relation exists. The results of this comparison are shown in Table 11.

Table 11. ASI relatedness of BC pairs.

Coupling strength of BC pairs	Percent of pairs related according to ASI
$n=5$ and over	86
$n=4$ and 3	71
$n=2$	67
$n=1$	55
100 random pairs ($n=0$)	0

The number of random pairs that can be formed from a population of N papers is $\frac{1}{2}(N^2 - N)$. In our case, $N = 334$ and the number of possible pairs is 55,611. Since the BC method produced 876 pairs, the probability of picking one such pair at random is less than two percent. We may, therefore, consider that for the one hundred random pairs n is virtually equal to zero.

Table 11 indicates that the coupling strength (n) is a measure of relatedness consistent with the judgment of ASI. The statement "Of the group of papers for which $n = 1$, 55 percent are related according to ASI" does not imply the corollary that the remaining 45 percent are not related. Of the remaining 45 percent, we can only say that they do not share an ASI category.

The comparison of Table 11 is asymmetric in its treatment of BC and ASI in the following sense. For BC we considered not only the fact of relatedness but also its relative strength. The judgment of ASI, on the other hand, was taken as a simple yes or no. Thus, we could count the number of yes and no's in each BC class of given coupling strength, but we could not in any given instance indicate a comparable ASI strength of coupling between any two papers. The results of the third stage comparison show that as the coupling strength (n) of the BC pairs increases, the probability of sharing an ASI category also increases.

• Fourth Stage

This experiment could not be performed because we found no consistent measure of strength of relatedness in the ASI method. Several attempts were made but all ended in failure. For example: Let $a, b, c, d \dots$ be various ASI categories. Now consider two papers both categorized as $P(a)$, are they more or less related than

two papers categorized $P(a,b)$ and $P(a,c)$? Or consider two papers, both categorized $P(a)$, are they more or less related than two papers, both categorized $P(a,b,c)$? The above and similar questions could not be answered in any consistent manner. This difficulty no doubt derives from the fact that the ASI categories are not of equal "logical size." We must bear in mind, however, that negative results are always tentative. We suspect that a close examination of the "logical size" of the ASI categories could yield a weighting factor that would give a measure of relatedness to the method.

• Summary

1. The papers in Vol. 112 of the *Physical Review* were processed by the method of Bibliographic Coupling and the results compared with those of Analytic Subject Indexing as performed by the editors of the *Physical Review*.

2. Groups formed by BC show high correlation with the verdict of the ASI method. Detailed results and numerical measures of this correlation are given.

3. Groups formed by ASI correlate with those formed by BC more or less, depending on the "logical size" of the ASI category.

4. The verdict of ASI confirms that the coupling strength of the BC method is a measure of strength of relatedness between pairs of papers. As the coupling strength goes up, the probability of relatedness by ASI also goes up.

5. No consistent measure of strength of relatedness between pairs of papers could be found applicable to the ASI method.

6. This report does not pass judgment on the utility of either method to any specific application.

References

1. M. M. Kessler, "An Experimental Study of Bibliographic Coupling Between Technical Papers," *IEEE Transaction PTGIT*, vol. IT-9, 1963, p. 49.
2. —, "Bibliographic Coupling Between Scientific Papers," *American Documentation*, vol. 14, no. 1, 1963, p. 10.
3. —, "Bibliographic Coupling Extended in Time: Ten Case Histories," *Inform. Stor. Retr.*, vol. 1, 1963, p. 109.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TECHNICAL INFORMATION PROGRAM*

SOME VERY GENERAL DESIGN CONSIDERATIONS

A discussion of the goals and constraints
that guided the evolution of the TIP system.

M. M. Kessler

October, 1967

*Supported by a grant from the National Science Foundation

SOME VERY GENERAL DESIGN CONSIDERATIONS

The evolution of the TIP system has been guided by certain general principles and considerations that we wish to discuss in this section.

We are concerned with the interaction between two populations, a) the population of documents and papers, each containing some form of scientific intelligence, and b) the population of readers, each in search of some form of scientific intelligence. The problem quite simply stated is to match, with high probability of success, the information content of one to the information needs of the other. This is not the same as matching answers to questions. In mechanizing the process of information search and retrieval, we should concentrate on the problems of finding likely sources of information. The result of the process must be a manageable list of papers whose probability of containing the information needed is very high compared to that of the general literature. Selection or rejection of specific items should then be left to the intelligence of the reader. The above is particularly true for the transfer of information between peripheral or widely separated fields as, say, between physics and biology. Information transfer gets more difficult as fields of specialization drift farther apart. Information flow between separated fields of investigation may be the critical problem area since we are faced with the seemingly contradictory trends of deeper specialization within each field and at the same time greater demands on complex systems that require cross-science communication.

It is quite possible that no single method or scheme of documentation and retrieval will satisfy all the various needs of the scientific community. Indeed, it is possible that existing schemes are already quite adequate and what is lacking is a flexible system that will allow all the various schemes to complement and reinforce each other and be responsive to the needs and desires of the user. The problems of compatibility of components, system organization, human engineering and sensitivity to user feed-back are perhaps as important as the ultimate efficiency of any one particular method of retrieval.

The system concept was so basic in the design of TIP that further elaboration may be useful. Consider by way of analogy a well-developed transportation system. It will be made up of many components: railroads, airplanes, automobiles, busses, rapid transit, etc. These may be combined in any one of many combinations to satisfy the needs of the user. It is the meshing and combination of all these components tied together by an efficient communication system that constitutes a proper transportation system. A given user will design his own mix of all the available facilities to suit his needs and convenience. It is in this sense that we have applied the system concept to our model information system. A large number of possibilities and operations are made possible. The proper mix of strategies and sequence of operations is left open and adjustable to the occasion.

A prototype system, in order to be useful, must satisfy two requirements. It must be capable of performing a real function in a realistic environment. This requires that it be above critical size in all of its components. Although to the designers of TIP the system is a subject for experimental observation, for the user, the system must be a reasonable approximation of the real thing. Only then will it be subject to proper utilization and test. This, to a large extent, determines the optimum or critical size of a model system. It must be large and complete enough to appear as a functioning tool to the user; on the other hand, it must not be so big as to offer great and costly resistance to change and modification. The system must, furthermore, be flexible enough to admit new components and devices to be incorporated into its existing structure. Above all, it must not be committed to any rigid a priori philosophy of operation. Clearly, these conditions may be mutually conflicting and can be only approximated. Nevertheless, the mere awareness of them offers a useful guide to system design.

A second important requirement for a prototype system is that it be capable of technological evolution and not constitute a dead end. Every component and process designed into a prototype system must be evaluated not only in terms of its immediate contribution to the model but also in terms of its ability to be scaled upward by a factor of at least ten. This latter

condition strongly affects the equipment and processing schemes that are admissible at this time. On the one hand, we exclude from our system all human input processing beyond the merest clerical manipulations on the grounds that expert judgment, evaluation, indexing, etc., require skills that are not ordinarily available in large numbers and are, at any rate, the very skills that our system must conserve and release for other purposes. At the other extreme, we eliminate, for the time being, schemes and processes that depend on computer manipulation of large bodies of text. Our system bases itself on a technological capability extrapolated to some three or five years from now (1967). The range of equipment capabilities at that time will not, in our present judgment, include an appreciable capacity to manipulate large bodies of text in digital form. Our system, therefore, does not anticipate this capability and excludes components that depend on it.

What we usually call "library automation" concerns the inventory control and manipulation of large lists of papers, books, documents or any other form of recorded data. Once human cataloging or indexing or abstracting has been accomplished, the remaining problems are not much different from those encountered in other services that dispense to the public. There is not much in "library automation" in this sense that differs appreciably from what is encountered in large banks and insurance or mail order companies, to say nothing of the Internal Revenue Service, Social Security and the Army Quartermaster Corps. The management of indexed lists, as important as it is, does not represent the main thrust and promise of computer technology in the field of information handling. The really revolutionary dream that some of us indulge in is that the computer, as a logical machine, has application not only as a bookkeeper, but more significantly as a correlator of recorded knowledge and a discoverer of relations between isolated bits of information whose record, although available, is so dispersed that the integrated pattern does not emerge. Knowledge, like energy, in order to be useful must be gathered into a concentrated package. It must be brought into a focus of space and time. Only then can human imagination play on it and acquire insight.

Scientific information is packaged and transmitted in various forms such as articles, books, reports, theses, etc. The most successful and ubiquitous carrier of scientific information, indeed the most characteristic in the entire history of science, is the journal article. The strength and flexibility of this form of scientific self-expression is obvious from its ability to serve equally well the routine and pedestrian results of some laboratory measurement as well as the imaginative output of a genius such as Einstein or Bohr. The "scientific paper" is such a marvelous and accommodating invention that we seem to take it for granted and forget that it has form and structure that fit its function. Regardless of any value judgment as to the worth of the scientific work that it records, the paper is designed to convey an operational experience. We must understand that a scientific paper is already a highly stylized and encoded message of the writer's operational experience. We regard the scientist and his operational behavior as the proper information; the paper or document is the message that carries the information. What is needed is an address that will deliver the message to the intended reader with a high probability of success.

We must not assume that the address is either a summary of the paper or a distilled and abbreviated portion of it. The address should be a separate message that tells us not what is in the paper but to whom it must be delivered. It is a directional vector, a command to deliver. Ideally, this vector should be determined from a knowledge of the author and his community of colleagues as well as of the prevailing conditions in science and technology. The latter point is of critical significance. The origin of the addressing vector is a unique experience of one man at a given time, namely the author of the paper. The terminus of the vector cannot be so uniquely described. The meanings, implications, and importance of a piece of scientific work depend not only on its contents but also on the particular condition of the scientific environment. Indeed, the paper acquires new meaning as it is used and interpreted by the community of scientists, even as a new law acquires meaning by precedent as it is used and interpreted by the courts. The standard indexing and cataloging techniques, even under the best of circumstances, carry the danger of freezing the literature

and its significance to one particular time in the evolution of science. What is equally serious is that the indexing of a paper is done from the point of view of one particular field or discipline. Thus, physics papers are indexed for physicists. Unless the paper is obviously of interest to biology at the time of indexing, the matter may be lost to biologists forever. A paper in The Physical Review may become obsolete to the readers of The Physical Review after a given period of time, but it may just then become of current interest and use to the readers of the chemical or biological literature. For these reasons, it is dangerous to index a paper once and for all and expect a viable retrieval system.

The safest way to generate the address of a paper is always to examine the operational experience of the author as well as the condition of the scientific environment at the given time. In most cases, the only available evidence of the author's experience is his encoded message unit, namely the paper or document. Therefore, in examining the paper for purposes of address at any given time, we must look for clues that refer to the author and his experience and avoid the properties of words and language that are incidental to the verbalization of the experience. The experience itself may be a truly creative effort or a mere compulsive and ritualistic act. Regardless of any value that the paper may possess at any given time, there is a formal pattern common to a great majority of papers. By far the largest number of scientific papers report the Results of a Man Operating on a System. Indeed, from our point of view we may equate the two and say that the scientific paper is the result of a man operating on a system. Since these four components are quite independent of each other and seem to be basic to the scientific act, we postulate that they may be used as four components of a directional vector that may constitute the address of a paper.

At the present time the almost universal mode of communicating scientific papers is to bunch them into groups called journals which are then delivered to a list of subscribers. This will produce rather loosely correlated groups of papers that depend not only on traditional notions of subject classification but also on the accidents of language, geography, and professional affiliations of

the author. We should probably continue to publish serial journals for archival and historical reasons. Nor should the proliferation of journal titles alarm us since it is an obvious reflection of the growth of scientific activity. But it is quite clear that we cannot continue to depend on a single mailing list as the proper address generator for all the papers in a given journal. A more reasonable approach would be to accumulate all the articles in a given set of journals and re-group them into more meaningful packages for purposes of address and communication. We are suggesting a finer bunching process that will generate a more differentiated and individualized set of "journals" to fit the needs of individual workers or groups. These "journals" need not be actually and physically produced in print on paper. They may exist as virtual or potential subsets of a large library. Indeed, ideally they need not even be generated in any a priori fashion but come into being as the result of an interaction between a user and an intelligence-handling machine. This possibility is not confined to the recent literature only. If we consider retrieval as communication with the accumulated literature of the past, similar considerations may apply.

It is unfortunate that so much attention is given to retrieval of information and so little to an intelligent and well addressed communication in the first place. Retrieval implies that information is unavoidably lost into an undifferentiated mass and must be retrieved; communication implies that there is hope for arranging well-delineated channels and methods of address, finer than journal subscription, that will direct information to proper receivers. In the real world, both processes, retrieval and communication, exist, and one should not consider one as excluding the other. The best system of communication will not eliminate the need of retrieval, although it may considerably lighten its load. On the other hand, no amount of retrieval can replace a communication system well-balanced in its various capabilities to offer both broadcast type of information of general use to large groups of people as well as more sharply delineated messages for specialized groups or even individual needs.

Our concept of the proper relation between the published literature, retrieval, communication, and the population of readers is shown in Fig. 1. In this chart, retrieval and selection are indicated as elements of the processing

scheme that the published literature must undergo prior to its proper dissemination to the scientific community. The processing function operates on the totality of the published literature and re-groups the papers into meaningful bundles which may then be addressed and delivered to the appropriate readers. The output from processing takes two channels, i.e. interrogated output when the address is an integral part of the interrogation, and non-interrogated output when the address is part of the processing function. The non-interrogated output may be broadcast or published through a variety of media to the community at large, or it may be transmitted to specific persons or groups. The interrogated output is normally available to the interrogator only, although it may at times be considered general enough to be worthy of broadcast, especially if the original interrogator adds to its value in some way by rearranging or otherwise enriching its content. The arrows leading from the bottom of Fig. 1 back to the input stage indicate a feed-back channel from the user.

The design of a system must take into account the probable or expected population of users. As a general rule, a system ought not to commit itself to a narrow or specialized user population. In our case, the design philosophy was to provide functional building blocks that can be grouped and re-grouped to fit individuals or groups of varying needs and backgrounds. A viable system must be capable of functioning as a source of current awareness as well as of information in depth. It must serve with equal facility the graduate student working on a specialized thesis problem and the mature scientist writing a general review or monograph. Indeed, the system ought to respond to the needs of the historian, sociologist, or manager of science as well as to the investigator. This goal can be achieved if the system is flexible enough to allow individual play and variation of basic search principles that allow each user to try various prescriptions or strategies and choose those which suit him best. One must by all means avoid a system design that is dedicated to a rigid mode and forces the user to adapt his needs to that mode. On the contrary, the system must be capable of change and adaptation, not only in response to its designers, but also in response to the daily manipulations by the users.